

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:24:08 ; Search time 32.13 Seconds  
(without alignments)  
437.057 Million cell updates/sec

Title: US-09-072-384-18  
Perfect score: 2080  
Sequence: 1 MAGIPGLFLFLFLCAVGQ.....LKYAQICYWIKGYLDREG 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.3	316	1	GSEP_BACLI
2	135.5	6.5	313	1	MPR_BACSU
3	124	6.0	260	1	ELI_BOVIN
4	121.5	5.8	490	1	FA10_RABIT
5	119	5.7	269	1	EL2_PIG
6	115.5	5.6	269	1	EL2A_HUMAN
7	115.5	5.6	271	1	EL2_MOUSE
8	115	5.5	376	1	FA10_TROCA
9	114.5	5.5	238	1	TRY5_AEDAE
10	113	5.4	266	1	EL1_PIG
11	113	5.4	266	1	EL1_RAT
12	112.5	5.4	259	1	TRYP_STRGR
13	112.5	5.4	273	1	YDGD_ECOLI
14	111.5	5.4	488	1	FA10_HUMAN
15	111.5	5.4	875	1	NETR_HUMAN
16	111	5.3	253	1	CFAD_HUMAN
17	111	5.3	761	1	NETR_MOUSE
18	110	5.3	269	1	EL2B_HUMAN
19	110	5.3	492	1	FA10_BOVIN
20	108	5.2	583	1	FA10_HUMAN
21	108	5.2	786	1	STUB_DROME
22	107.5	5.2	686	1	MA52_HUMAN
23	107	5.1	603	1	CFAI_MOUSE
24	105.5	5.1	604	1	CFAI_RAT
25	105	5.0	269	1	EL2_BOVIN
26	105	5.0	274	1	TRY5_ANOGA
27	104.5	5.0	268	1	TRYP_STRGA
28	103.5	5.0	271	1	EL2_RAT
29	103.5	5.0	416	1	FA9_BOVIN
30	101.5	4.9	281	1	TRY2_DROER
31	99	4.8	855	1	ST14_MOUSE
32	98.5	4.7	260	1	COGS_HYPLI
33	98.5	4.7	267	1	TRY7_ANOGA

## RESULT 1

ID	GSEP_BACLI	STANDARD;	PRT;	316 AA.
AC	P80057;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC			
DE	ENDOPEPTIDASE) (GSE).			
GN	BLASE.			
OS	Bacillus licheniformis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1402;			
RN	[1]			
RC	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RP	STRAIN=ATCC 14580;			
RX	MEDLINE=93054737; PubMed=1429718;			
RA	Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,			
RA	Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;			
RT	"Purification, characterization, cloning, and expression of a			
RT	glutamic acid-specific protease from Bacillus licheniformis ATCC			
RT	14580.";			
RL	J. Biol. Chem. 267:23782-23788(1992).			
RN	[2]			
RP	SEQUENCE OF 95*316.			
RX	MEDLINE=92155199; PubMed=1346764;			
RA	Svensden I., Bredam K.;			
RT	"Isolation and amino acid sequence of a glutamic acid specific			
RT	endopeptidase from bacillus licheniformis.";			
RL	Eur. J. Biochem. 204:165-171(1992).			
CC	- - FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE			
CC	CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG			
CC	PREFERENCE FOR GLU.			
CC	- - CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP- -XAA, GLU- -XAA.			
CC	- - SUBCELLULAR LOCATION: SECRETED.			
CC	- - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE			
CC	V8 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D10060; BAA00949.1; -			
DR	PIR; S23078; S23078.			
DR	PIR; A45134; A45134.			
DR	MEROPS; S01.271; -			
DR	InterPro; IPR00126; Ser.proteas_v8.			
DR	InterPro; IPR001254; Trypsin.			
DR	Pfam; PF00089; trypsin; 1			
DR	PRINTS; PR00839; V8PROTEASE.			
DR	SMART; SM00020; Tryp_SPC; 1.			

## ALIGNMENTS

P19540 canis famil  
P28175 tachypleus  
P52905 drosophila  
P49275 dermatophag  
P16292 oryctolagus  
P57727 homo sapien  
Q01513 podospora a  
P35035 anopheles g  
Q975Y6 homo sapien  
P03049 drosophila  
Q28422 carcinoscor  
P39675 dermatophag

34 97.5 4.7 452 1 FA9\_CANFA  
35 97 4.7 1019 1 LFC\_TACTR  
36 96.5 4.6 252 1 TRY1\_DROME  
37 96.5 4.6 259 1 DEF3\_DERFA  
38 96.5 4.6 275 1 FA9\_RABIT  
39 96.5 4.6 454 1 TMS3\_HUMAN  
40 96.5 4.6 2145 1 CYAA\_PODAN  
41 96 4.6 274 1 TRY1\_ANOGA  
42 96 4.6 855 1 ST14\_HUMAN  
43 95.5 4.6 430 1 SNAK\_DROME  
44 95 4.6 1019 1 LFC\_CARRO  
45 94.5 4.5 261 1 DER3\_DERPT

```

RX MEDLINE=89108019; PubMed=3145906;
RA Smith H., de Jong A., Bron S., Venema G.;
RT "Characterization of signal-sequence-coding regions selected from the
RL Bacillus subtilis chromosome.";
RN Gene 70:351-361(1988).
[4]
RP SEQUENCE OF 1-10 FROM N.A.
RC STRAIN=168;
RX MEDLINE=96118702; PubMed=7496533;
RA Saxild H.H., Jacobsen J.H., Nygaard P.;
RT "Functional analysis of the Bacillus subtilis purT gene encoding
RL formate-dependent glycylamide ribonucleotide transformylase.";
RN Microbiology 141:2211-2218(1995).
CC -1- SUBUNIT: MONOMER
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
CC V8 FAMILY.
CC -1- CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2
CC FAMILY OF SERINE PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L140505; AAA22604.1; .
DR EMBL; AB006424; BAA33121.1; .
DR EMBL; M22916; AAA22832.1; .
DR EMBL; Z99105; CAB12018.1; .
DR PIR; A35122; A35122.
DR HSP; P00756; LSGF.
DR MEROPS; S01.272; .
DR Subtilist; BG10690; mpr.
DR InterPro; IPR000126; Ser_proteas_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00020; Ttyp_SPC; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
DR Hydrolyase; Serine protease; Signal; Zymogen; Complete proteome.
KW SIGNAL 1 34 POTENTIAL.
FT PROPEP 35 93
FT CHAIN 94 313 EXTRACELLULAR METALLOPROTEASE.
FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 131 147 BY SIMILARITY.
FT CONFLICT 61 68 QVSAPVEG -> PLESTAQA (IN REF. 3).
FT SEQUENCE 313 AA; 33842 MW; D41788E8D652AE94 CRC64;
SQ
Query Match 6.5%; Score 135.5; DB 1; Length 313;
Best Local Similarity 22.2%; Pred. No. 0.00035;
Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;
QY 72 GTPLPYEEAKQYLSYETLYANGSRTRTQVGIVILSSGD--GAQHRDSSGSKRRKQ 129
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
29 GVPKAAENPQTSVNSGTKEADATKNQT-----SKADQVSAPYECTGKTSKS----- 75
QY 130 IYGYSRF-----SIFGKD-----FLLNYPSTSVKLST-----GCTGTL 164
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
76 LYGGQTELEKNIQTLPSSIIIGTDERTRISSTTSFPFRTVLSIKYVNTSSYVCTGFL 135
QY 165 VAEKHVLTAARHIIH-----DGRTYVKGTQKLRVGLKPKFKDGGRGAND 208
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
136 VNPTVVTVTACHCVYSQDHGWNASTITAAPGRGSSYPVGTV----- 175
QY 209 STSAMPQMFQWIRVKRTHVPKWTGN---ANDIGMDYDYLLELKKPKRKFMKIGV 265
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
176 -SGTMFYSVK-GWTESKDTNDYGAIKLNGSPGNTVGW-YGYRTTNSSP-----VGL 225

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:17:56 ; Search time 48.03 Seconds  
(without alignments)  
607.430 Million cell updates/sec

Title: US-09-072-384-18  
Perfect score: 2080  
Sequence: 1 MAGIPGLLLFLFLCAVQ.....LKVAQICYWKNLYDCREG 383

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.3	316	2 A45134	endopeptidase (EC
2	135.5	6.5	313	2 A35122	metalloproteinase
3	120.5	5.8	482	1 EXPT	coagulation factor
4	119	5.7	269	2 A26823	pancreatic elastase
5	117	5.6	522	2 T29767	hypothetical prote
6	115.5	5.6	269	2 B26823	pancreatic elastase
7	115.5	5.6	271	2 A25528	pancreatic elastase
8	115	5.5	258	2 S70439	pancreatic elastase
9	115	5.5	267	4 A56615	probable pancreati
10	114.5	5.5	238	1 TRKV5Y	trypsin-like prote
11	114	5.5	246	1 DBHU	complement factor
12	113	5.4	266	1 ELRT1	pancreatic elastase
13	113	5.4	266	1 ELPG	pancreatic elastase
14	112.5	5.4	259	1 TRSMG	trypsin (EC 3.4.21
15	112.5	5.4	273	2 E85765	hypothetical prote
16	112.5	5.4	273	2 H64915	hypothetical prote
17	111.5	5.4	488	1 EXHU	coagulation factor
18	111	5.3	761	2 JC5759	brain-specific ser
19	110	5.3	269	2 C26823	pancreatic elastase
20	110	5.3	492	1 EXBO	coagulation factor
21	110	5.3	1582	2 T15308	hypothetical prote
22	109.5	5.3	405	2 T35117	probable secreted
23	108	5.2	583	2 A23154	complement factor
24	108	5.2	786	1 A47547	serine proteinase
25	108	5.2	1047	2 A55617	masquerade precurs
26	107.5	5.2	236	2 A28566	T-cell suppressor
27	107.5	5.2	686	1 A59271	Ra-reactive factor
28	105	5.0	274	2 SA0004	trypsin-related pr
29	103.5	5.0	271	1 ELRT2	pancreatic elastase

30	103.5	5.0	416	1 KFBO	coagulation factor
31	102.5	4.9	1238	2 T34929	hypothetical prote
32	100	4.8	272	2 JC4170	trypsin-like prote
33	98.5	4.7	267	2 S40006	trypsin (EC 3.4.21
34	98	4.7	548	2 D82175	probable trypsin V
35	98	4.7	624	2 T02289	probable polyalac
36	97.5	4.7	409	2 T35118	probable secreted
37	97.5	4.7	452	1 A30351	coagulation factor
38	97.5	4.7	747	2 I51579	complement factor
39	97	4.7	1019	2 A38738	coagulation factor
40	96.5	4.6	259	2 S68424	allergen Der f III
41	96.5	4.6	275	2 I46712	factor IX - rabbit
42	96.5	4.6	2145	2 JC4747	adenylate cyclase
43	95.5	4.6	430	1 A24702	serine proteinase
44	95	4.6	782	2 T32155	hypothetical prote
45	94	4.5	268	2 S68825	pancreatic elastase

ALIGNMENTS

RESULT 1

A45134  
endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis  
C:Species: Bacillus licheniformis  
C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999  
C:Accession: A45134; S23078  
R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S  
J. Biol. Chem. 267, 23782-23788, 1992  
A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s  
A:Reference number: A45134; MUID:93054737  
A:Accession: A45134  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-316 <KAK>  
A:Cross-references: GB:D10060; NID:9216263; PIDN:BAA00949.1; PID:d1001415; PID:g21626  
A:Experimental source: ATCC 14580  
A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIP:118785)  
R: Svendsen, I.; Bredem, K.  
Eur. J. Biochem. 204, 165-171, 1992  
A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase  
A:Reference number: S23078; MUID:92155199  
A:Accession: S23078  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 95-316 <SVR>  
C:Keywords: hydrolase

Query Match 9.3%; Score 194; DB 2; Length 316;

Best Local Similarity 24.0%; Pred. No. 1.2e-08;

Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;

QY	63	SSCGPOCHKGTPL---PTVEAKQYLSYETLYANGSRTETQVGIYLLSSSGCAQHRDSC	119
DB	28	AQAAPSPH--TPVSDPSY-KAETSVYDP-----NIKSDQYGLYKAFGTGKVNETKE	79
QY	120	SSGKSRKRQIYGYDGRSTFGKDFLLN-----YPFSTSVKLSL---GCTGLVAEKHV	170
DB	80	KAEKSPAKAPY---SIKSVIGSDDRVTNTTAYPRATVHSSIGSGCTGMIGPKTV	136
QY	171	LTAACHIDPKGT-YVKGTOKLRVFLKPKFKDGGGRANDSTAMPQMKPQWIRVRKTHV	229
DB	137	ATAGHCYDTSSGSFAGTATVSPG-----RNCTS-----YPVGSVKSTRYFI	178
QY	230	PKGWIKGNANDICMDYDYLLELKKPKHKFKMIGVSPPAKQLPGCRIHFSGYDNDRPGN	289
DB	179	PSGRSGNTN-----YDYGAIELSEPIGNTVGVFGYSYTTSSLVGTTVTISGYPGDKTAG	233
QY	290	LVRFCDD--VKDETLLYQQCDQAQPGASGVYVRMKRQOQKWERKII---GIFSGH	343
DB	234	TQWHSGPATSEITYKLYQYAM-DTYGQSGSPVFESSSRNTCSGPCSLAVHTNGVYGG-	291

QY 344 QWDMNGSPQDNVAVRITPLKVAQICW 372  
Db 292 -----SSYNGRTRITKEVFDNLTNW 311

RESULT 2  
Metalloproteinase (EC 3.4.21.6) mpr precursor, extracellular - Bacillus subtilis  
A:Species: Bacillus subtilis  
C:Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 15-Oct-1999  
C:Accession: A35122; I40010; A69660  
R:Sloma, A.; Rudolph, C.F.; Rufo Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; P.  
J. Bacteriol. 172, 1024-1029, 1990  
A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.  
A:Reference number: A35122; MUID:90130256  
A:Accession: A35122  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-313 <SLO>  
A:CROSS-references: GB:L10505; NID:g143209; PIDN:AAA22604.1; PID:g143210; GB:M29036  
R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.  
Gene 70, 351-361, 1988  
A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s  
A:Reference number: I39994; MUID:89108019  
A:Accession: I40010  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-60,65,'L',67,'S',69,'AOA' <RES>  
A:CROSS-references: GB:M22916; NID:g143701; PIDN:AAA22832.1; PID:g143702  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berten  
C.; Bron, S.; Brouillet, S.; Bruchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, V.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroette, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69660  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-313 <KUN>  
A:CROSS-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12018.1; PID:ell82176.  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: mpr  
C:Keywords: hydrolase

Query Match 6.5%; Score 135.5; DB 2; Length 313;  
Best Local Similarity 22.2%; Pred. No. 0.00091;  
Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;

QY 72 GTPLPTVEAKQYLSYETLVANGSRRTQVGIYLLSSGD--GAQHRDSCGSKSRKRQ 129  
Db 29 GVPAAKAENPOTSNSGKKEADATKQNT-----SKADQVSAPYEGTGTKTS----- 75

QY 130 IYGVDSRF-----SIFGKD-----FLNYPSTSVKLEST-----GCTGTL 164  
Db 76 LYGGQTELEKNIQLOTPSSIIIGTDERTRISTTSFPRAIVQLSIKYPNVSSTYGCCTGFL 135

QY 165 VAEKHVLTAAHCIIH-----DGKTYVKGQKLRFLKPKFKDGGRGAND 208  
Db 136 VNPNTVTVAGHCYVSDHQWASTITAAAPGRNGSSYPYGTV----- 175

QY 209 STSAMPQMFQWTRVRKTRVTHVPGKWIKN---ANDICWDYDYLLELKKPKRKFEMKIGV 265

Db 176 -SGTMFYSVK-GWTESKDTNYDYGAIKLNGSPGNTVGW-YGYRTTNSSP-----VGL 225  
QY 266 SPPAKQLPGGRHFGSDYNDPRGNLVYFCDKDYDLYLLYQCDQAQSGSGVYVWM 325  
Db 226 SSSVTGPFCDKTFGTMTSDTKPIR-----SAETYLKLT-TTDTYGCQSGSPVY----- 272

QY 326 KROQKWERKIIIGFSGHQWDMNGSPQDNVAVRITPLKVAQICW 372  
Db 273 -RNYSDTCQTALHT-----NGG-SSYNLGTRVNDVFNNTQYW 310

RESULT 3  
EXRT  
coagulation factor Xa (EC 3.4.21.6) precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Jan-1995 #sequence\_revision 07-Feb-1997 #text\_change 08-Dec-2000  
C:Accession: S49075; JC4670; PS0191; PS0190; I62745  
R:Stanton, C.; Ross, P.; Hutson, S.; Wallin, R.  
Thromb. Res. 80, 63-73, 1995  
A:Title: Evidence for competition between vitamin K-dependent clotting factors  
A:Reference number: A58498; MUID:96093366  
A:Accession: S49075  
A:Molecule type: mRNA  
A:Residues: 1-482 <STAL>  
A:CROSS-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
A:Note: submitted to the EMBL Data Library, June 1994  
A:Note: neither the complete nucleic acid sequence nor the complete translation are s  
R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.  
Gene 189, 269-273, 1996  
A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.  
A:Reference number: JC4670; MUID:96194815  
A:Accession: JC4670  
A:Molecule type: mRNA  
A:Residues: 1-482 <STPA2>  
A:CROSS-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
A:Experimental source: Cos-1 cell  
R:Enjoji, K.; Miyazaki, K.; Kato, H.  
J. Biochem. 109, 890-898, 1991  
A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat  
A:Reference number: PS0190; MUID:92041742  
A:Accession: PS0191  
A:Molecule type: protein  
A:Residues: 41-58,'X',60-65 <ENJ1>  
A:Accession: PS0190  
A:Molecule type: protein  
A:Residues: 183-186,'X',188-207 <ENJ2>  
R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.  
Eur. J. Haematol. 52, 162-168, 1994  
A:Title: Analysis of the partial nucleotide sequences and deduced primary struct  
A:Reference number: I46196; MUID:94222160  
A:Accession: I62745  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 295-383,'G',385-455 <MUR>  
A:CROSS-references: GB:D21215; NID:g415309; PIDN:BAA04756.1; PID:g455396  
C:Function:  
A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the  
A:Pathway: blood coagulation  
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol  
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-40/Domain: propeptide #status predicted <PRO>  
F:25-84/Domain: Gla domain homology <GLA>  
F:41-179/Product: coagulation factor X light chain #status predicted <LCH>  
F:90-121/Domain: EGF homology <EG1>  
F:129-164/Domain: EGF homology <EG2>  
F:183-482/Product: coagulation factor X heavy chain #status predicted <HC>  
F:183-231/Domain: activation peptide #status predicted <APT>  
F:232-482/Product: coagulation factor Xa heavy chain #status predicted <ACT>  
F:232-460/Domain: trypsin homology <TRY>  
F:45,47,54,56,59,60,65,66,69,72,79/Modified site: gamma-carboxyglutamic acid (Glu) #s  
F:57-62,90-101,95-110,112-121,129-140,136-149,151-164,172-340,238-243,259-275,388-402

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:25:48 ; Search time 84.84 Seconds  
(without alignments)  
675.846 Million cell updates/sec

Title: US-09-072-384-15  
Perfect score: 2131  
Sequence: 1 MAGIPGLFLFLFLCAVQ.....IKGYLDCREGDTVPFPGSN 392

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	2080	97.6	383	4	Q95084
2	1884.5	88.4	382	11	Q9D6X6
3	1042	48.9	413	4	Q9BQP6
4	133.5	6.3	799	11	Q9DBI0
5	129.5	6.1	303	2	Q9EXR9
6	129	6.1	469	6	Q9GMD9
7	127.5	6.0	1322	5	Q9NAT0
8	126.5	5.9	678	11	Q9JJJ58
9	124.5	5.8	1322	5	Q9NJS5
10	124	5.8	266	6	O46644
11	123.5	5.8	339	11	Q9QX91
12	123.5	5.8	366	11	Q9QX85
13	123.5	5.8	541	11	Q9QX90
14	123.5	5.8	623	11	Q9JJJ3
15	123.5	5.8	643	11	Q9QX84
16	122	5.7	259	5	Q9XV61
17	122	5.7	449	5	Q9VDJ8
18	120.5	5.7	482	11	Q63207
19	118	5.5	266	11	Q9D936

20	118	5.5	573	5	Q9V516
21	118	5.5	1376	5	Q9VQR8
22	117	5.5	258	5	Q9W5U8
23	117	5.5	522	5	O01771
24	115.5	5.4	269	4	Q14243
25	115	5.4	258	4	Q9UN11
26	114.5	5.4	279	5	Q9G991
27	113.5	5.3	418	5	Q9VAB7
28	113.5	5.3	685	11	Q92338
29	113.5	5.3	737	13	Q90422
30	112.5	5.3	603	5	Q9VDV1
31	112.5	5.3	686	4	Q9BZHO
32	112	5.3	249	13	Q9W7Q1
33	111.5	5.2	481	11	O54740
34	110	5.2	1449	5	Q10922
35	109.5	5.1	405	2	O69973
36	108.5	5.1	481	11	O88947
37	108.5	5.1	481	11	O99L32
38	108	5.1	274	5	Q17086
39	108	5.1	750	13	Q9W633
40	108	5.1	1047	5	Q24019
41	107.5	5.0	236	11	Q9Z1H1
42	107.5	5.0	427	10	Q9AYR4
43	107.5	5.0	611	5	Q9VZS8
44	106	5.0	460	5	Q9VFZ6
45	106	5.0	762	13	Q9YIC6

## ALIGNMENTS

RESULT	1
Q95084	
ID	Q95084
AC	Q95084; PRELIMINARY; PRT; 383 AA.
DT	01-MAY-1999 (Tremblrel. 10, Created)
DT	01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE	SERINE PROTEASE (HYPOTHETICAL 43.0 KDA PROTEIN) (PROTEASE, SERINE, 23).
DE	23).
GN	ZSGL3 OR DKFZP586B0719.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=UMBILICAL VEIN;
RA	Li X., Tedder T.F.;
RT	"A novel serine protease from human umbilical vein endothelial cells".
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RA	Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.;
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	TISSUE=UTERUS;
RA	Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN	[4]
RP	SEQUENCE FROM N.A.
RC	TISSUE=CERVIX CARCINOMA;
RA	Strausberg R.;
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC	-1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC	-1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR	EMBL; AF015287; AAD01553.1;
DR	EMBL; AF193611; AAF07186.1;
DR	EMBL; AF136914; CAB56848.1;
DR	EMBL; BC001278; AA01278.1;
DR	MEROPS; S01.309; -.

```
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 383 AA; 43001 MW; 46E86C11ABFD58F CRC64;

Query Match          97.6%; Score 2080; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 1.8e-182;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFFLLCAVGQVSPYSAFWKPTWPAYRLPVVLPQSTLNLAKEHVLTAHCHIDG 60
DB 1 MAGIPGLLFLFFLLCAVGQVSPYSAFWKPTWPAYRLPVVLPQSTLNLAKEHVLTAHCHIDG 60
QY 61 VSSSCGQCHKGKPTLPYEEAKQYLSYETLYANGSRRTETOVGIYILSSSGDGAQHRDGS 120
DB 61 VSSSCGQCHKGKPTLPYEEAKQYLSYETLYANGSRRTETOVGIYILSSSGDGAQHRDGS 120
QY 121 SGKRRRQIYGVDSRFSIFGKDFLLNPFSTSVKLTGCTGLVAEKHVLTAHCHIDG 180
DB 121 SGKRRRQIYGVDSRFSIFGKDFLLNPFSTSVKLTGCTGLVAEKHVLTAHCHIDG 180
QY 181 KTYVKGTKLRVGLPKPKFKDGGGANDSTSAMPEQMKFQWIRKRVKTHVPKGIKGNAND 240
DB 181 KTYVKGTKLRVGLPKPKFKDGGGANDSTSAMPEQMKFQWIRKRVKTHVPKGIKGNAND 240
QY 241 IGMDDYALLELKKPKHKKRPMKIGVSPPAKQLPGRIHFGSYDNDRPNLVYRCDVKDE 300
DB 241 IGMDDYALLELKKPKHKKRPMKIGVSPPAKQLPGRIHFGSYDNDRPNLVYRCDVKDE 300
QY 301 TYDLLYQCCDAQPGASGSGYVVRMKRQOKWERRKIIGIFSGHGWDMNGSPQDFNVAVR 360
DB 301 TYDLLYQCCDAQPGASGSGYVVRMKRQOKWERRKIIGIFSGHGWDMNGSPQDFNVAVR 360
QY 361 ITPLKYAQICYWIKGNLYDCREG 383
DB 361 ITPLKYAQICYWIKGNLYDCREG 383

RESULT 2
Q9D6X6 PRELIMINARY; PRT; 382 AA.
ID Q9D6X6 PRELIMINARY; PRT; 382 AA.
AC Q9D6X6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 2310046G15RIK PROTEIN.
GN 2310046G15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
```

```
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AK009847; BAB26541.1; -.
DR MGD; MGI:1923703; 2310046G15RIK.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 382 AA; 43053 MW; 69C9A7080E5B2306 CRC64;

Query Match          88.4%; Score 1884.5; DB 11; Length 382;
Best Local Similarity 90.6%; Pred. No. 1.5e-164;
Matches 347; Conservative 15; Mismatches 20; Indels 1; Gaps 1;

QY 1 MAGIPGLLFLFFLLCAVGQVSPYSAFWKPTWPAYRLPVVLPQSTLNLAKEHVLTAHCHIDG 60
DB 1 MAGIPGLLFLFFLLCAVGQVSPYSAFWKPTWPAYRLPVVLPQSTLNLAKEHVLTAHCHIDG 59
QY 61 VSSSCGQCHKGKPTLPYEEAKQYLSYETLYANGSRRTETOVGIYILSSSGDGAQHRDGS 120
DB 61 VSSSCGQCHKGKPTLPYEEAKQYLSYETLYANGSRRTETOVGIYILSSSGDGAQHRDGS 119
QY 121 SGKRRRQIYGVDSRFSIFGKDFLLNPFSTSVKLTGCTGLVAEKHVLTAHCHIDG 180
DB 121 SGKRRRQIYGVDSRFSIFGKDFLLNPFSTSVKLTGCTGLVAEKHVLTAHCHIDG 179
QY 181 KTYVKGTKLRVGLPKPKFKDGGGANDSTSAMPEQMKFQWIRKRVKTHVPKGIKGNAND 240
DB 181 KTYVKGTKLRVGLPKPKFKDGGGANDSTSAMPEQMKFQWIRKRVKTHVPKGIKGNAND 239
QY 241 IGMDDYALLELKKPKHKKRPMKIGVSPPAKQLPGRIHFGSYDNDRPNLVYRCDVKDE 300
DB 241 IGMDDYALLELKKPKHKKRPMKIGVSPPAKQLPGRIHFGSYDNDRPNLVYRCDVKDE 299
QY 301 TYDLLYQCCDAQPGASGSGYVVRMKRQOKWERRKIIGIFSGHGWDMNGSPQDFNVAVR 360
DB 301 TYDLLYQCCDAQPGASGSGYVVRMKRQOKWERRKIIGIFSGHGWDMNGSPQDFNVAVR 359
QY 361 ITPLKYAQICYWIKGNLYDCREG 383
DB 361 ITPLKYAQICYWIKGNLYDCREG 382

RESULT 3
Q9BQP6 PRELIMINARY; PRT; 413 AA.
ID Q9BQP6 PRELIMINARY; PRT; 413 AA.
AC Q9BQP6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DJ223E3.1 (PUTATIVE SECRETED PROTEIN ZS1G13).
GN DJ223E3.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RX EMBL: AL121939; CA35071.1; -.
RA Dunn M.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121939; CA35071.1; -.
SQ SEQUENCE 413 AA; 47098 MW; 818D9C951BD2D6C1 CRC64;
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:16:53 ; Search time 64.47 Seconds  
(without alignments)  
450.391 Million cell updates/sec

Title: US-09-072-384-15  
Perfect score: 2131  
Sequence: 1 MAGIPGLFLFLFLLCAVCG.....IKGNLYDCRGDTVPFPGSN 392

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_ll01:\*

1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2131	100.0	392	22	Human Zsig13 varia
2	2085	97.8	392	22	Human Zsig13 varia
3	2080	97.6	383	20	Human transmembran
4	2080	97.6	383	20	WC9927094 Seq ID 1
5	2080	97.6	383	20	Amino acid sequenc
6	2080	97.6	383	21	Protein encoded by
7	2080	97.6	383	21	Protein encoded by
8	2080	97.6	383	21	Human TANGO 186 pr
9	2080	97.6	383	21	Human signal pepti
10	2080	97.6	383	21	A bone marrow secr
11	2080	97.6	383	22	Human PRO307 prote

12	2080	97.6	383	22	Human Zsig13 varia
13	2073	97.3	383	22	Human membrane or
14	1909.5	89.6	375	21	Protein encoded by
15	1909.5	89.6	413	21	Protein encoded by
16	1890.5	88.7	382	21	Murine TANGO 186 p
17	1041	48.9	413	20	Human PRO1057 prot
18	1041	48.9	413	20	Human nn320_2 secr
19	1041	48.9	413	21	Human PRO1057 (UNQ
20	202.5	9.5	314	22	Bacillus lichenifo
21	194	9.1	316	13	Bacillus lichenifo
22	193	9.1	316	22	Protease BLase. S
23	184	8.6	222	22	Bacillus lichenifo
24	183	8.6	222	22	Bacillus lichenifo
25	183	8.6	222	22	Bacillus lichenifo
26	183	8.6	222	22	Bacillus lichenifo
27	183	8.6	222	22	Bacillus lichenifo
28	183	8.6	222	22	Bacillus lichenifo
29	181	8.5	222	22	Bacillus lichenifo
30	181	8.5	222	22	Bacillus lichenifo
31	181	8.5	222	22	Bacillus lichenifo
32	181	8.5	222	22	Bacillus lichenifo
33	180	8.4	222	22	Bacillus lichenifo
34	180	8.4	222	22	Bacillus lichenifo
35	180	8.4	222	22	Bacillus lichenifo
36	180	8.4	222	22	Bacillus lichenifo
37	179	8.4	222	22	Bacillus lichenifo
38	179	8.4	222	22	Bacillus lichenifo
39	179	8.4	222	22	Bacillus lichenifo
40	179	8.4	222	22	Bacillus lichenifo
41	179	8.4	222	22	Bacillus lichenifo
42	179	8.4	222	22	Bacillus lichenifo
43	179	8.4	222	22	Bacillus lichenifo
44	179	8.4	222	22	Bacillus lichenifo
45	179	8.4	222	22	Bacillus lichenifo

ALIGNMENTS

RESULT 1	
AAB48973	
ID	AAB48973 standard; Protein; 392 AA.
AC	AAB48973;
XX	
DT	27-MAR-2001 (first entry)
XX	
DE	Human Zsig13 variant #2, SEQ ID NO:15.
XX	
KW	Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue;
KW	glutaryl endopeptidase homologue; factor X homologue; trypsin homologue;
KW	trypsinogen homologue; mast cell protease homologue;
KW	collagenase homologue; protein degradation; food processing; brewing;
XX	alcohol production; laundry detergent component.
OS	Homo sapiens.
PN	US6153420-A.
PD	28-NOV-2000.
XX	
PF	04-MAY-1998; 98US-0072384.
XX	
PR	24-APR-1997; 97US-0044185.
PR	17-APR-1998; 98US-0062142.
XX	
PA	(ZYMO ) ZYMOGENETICS INC.
XX	
PI	Sheppard PO;
DR	WPI; 2001-060090/07.
XX	N-PSDB; AAC91783.

PT New isolated serine protease (designated zsig13), useful in industrial  
PT processes to degrade unwanted proteins or alter the characteristics of  
PT protein-containing composition, as well as in industrial applications  
PT (e.g. brewing)  
XX

PS Claim 1: Column 35-38; 26pp; English.

XX The invention relates to human zsig13 proteins (AAB48972-B48974), and  
CC to DNA encoding them (AAC91782-C91784). The invention also relates to  
CC expression vectors and host cells comprising a human zsig13 DNA, and the  
CC recombinant production of a human zsig13 protein or its precursor.  
CC zsig13 is a serine protease, and has significant homology to Bacillus  
CC licheniformis glutamyl endopeptidase, human clotting factor X, human  
CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine  
CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human  
CC zsig13 is located on chromosome 11q22.1. zsig13 is useful in industrial  
CC processes to degrade unwanted proteins or alter the characteristics of  
CC protein-containing compositions. It may also be used in industrial  
CC applications in which proteases are utilised, including food processing,  
CC brewing and alcohol production, and as a component of a laundry  
CC detergent. The present sequence represents a human zsig13 variant.  
XX

SQ Sequence 392 AA;

Query Match 100.0%; Score 2131; DB 22; Length 392;  
Best Local Similarity 100.0%; Pred. No. 3.8e-152;  
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLLFFLLCAGGVSPYSAPWKTWPAYRLPVVLPDSTLNLAKEPFGAEAKLE 60

DB 1 magipglilflilcavgvspysapwktwpayrlpvvlpdstlnlakpdfgaeakle 60

QY 61 VSSSCGPOCHKGTPLPYEAKOYLSVETLYANGSRTETOVGIYILSSSGDGAQHRDSDS 120

DB 61 vssscgpcqchkgtcplpyeakylsyetlyangstetqvgiylsssgdgqhrdsdgs 120

QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNPFSTSVKLSGTCTGLTVAEKHVLTAACHIDHG 180

DB 121 sgksrrkrqiygydsrfsifgkdfllnypfstsvklsstgctglvaekhvltaachidhg 180

QY 181 KTYVKGTKLRVGFLLPKFKDGGRGANDSTAMPQKQFOWIRKTHVPGKWIKNAND 240

DB 181 ktyvkgtklrvgflpkfkdggrgandstampsqkqfowirvkrthvpgkwikgnand 240

QY 241 IGMDDYVALLLEKPKHKKRPMKIGVSPPAKQLPGGRHFHSGYDNDRPGNLVYRFCDVKDE 300

DB 241 igmddyvallelkpkhkrkfmkigvspapakqlpggrihfsgydnndrpgnlvyrfcdvkde 300

QY 301 TYDLLYQCCDAQPGASGSGVYVVRMWRKQOKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360

DB 301 tydillyqcdagpgasggyvvyvmwrkrqqkwerkiigifsghqwvdmngspqdfnvavr 360

QY 361 ITPKVAQICYWIKGNVLDCREGDTVFPFGSN 392

DB 361 itplkvaqicywikgnvldcregdtvfpfgsn 392

RESULT 2

AAB48972

ID AAB48972 standard; Protein: 392 AA.

XX

AC AAB48972;

XX

DT 27-MAR-2001 (first entry)

DE

XX Human zsig13 variant #1, SEQ ID NO:2.

KW Human zsig13; serine protease; chromosome 11q22.1; elastase homologue;  
KW glutamyl endopeptidase homologue; factor X homologue; trypsin homologue;  
KW trypsinogen homologue; mast cell protease homologue;  
KW collagenase homologue; protein degradation; food processing; brewing;  
KW alcohol production; laundry detergent component.

XX Homo sapiens.

OS US6153420-A.

PN

XX 28-NOV-2000.

PD

XX 04-MAY-1998; 98US-0072384.

XX

PF 24-APR-1997; 97US-0044185.

XX

PR 17-APR-1998; 98US-0062142.

XX

PA (ZYMO ) ZYMOGENETICS INC.

XX

PI Sheppard PO.

XX

XX WPI; 2001-060090/07.

DR N-PSDB; AAC91782.

XX

XX New isolated serine protease (designated zsig13), useful in industrial

PT processes to degrade unwanted proteins or alter the characteristics of

PT protein-containing composition, as well as in industrial applications

PT (e.g. brewing)

PT

XX

XX Claim 1: Column 25-28; 26pp; English.

XX

XX The invention relates to human zsig13 proteins (AAB48972-B48974), and

CC to DNA encoding them (AAC91782-C91784). The invention also relates to

CC expression vectors and host cells comprising a human zsig13 DNA, and the

CC recombinant production of a human zsig13 protein or its precursor.

CC zsig13 is a serine protease, and has significant homology to Bacillus

CC licheniformis glutamyl endopeptidase, human clotting factor X, human

CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine

CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human

CC zsig13 is located on chromosome 11q22.1. zsig13 is useful in industrial

CC processes to degrade unwanted proteins or alter the characteristics of

CC protein-containing compositions. It may also be used in industrial

CC applications in which proteases are utilised, including food processing,

CC brewing and alcohol production, and as a component of a laundry

CC detergent. The present sequence represents a human zsig13 variant.

XX

XX Sequence 392 AA;

SQ

Query Match 97.8%; Score 2085; DB 22; Length 392;

Best Local Similarity 98.2%; Pred. No. 1.1e-148;

Matches 385; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLLFFLLCAGGVSPYSAPWKTWPAYRLPVVLPDSTLNLAKEPFGAEAKLE 60

DB 1 magipglilflilcavgvspysapwktwpayrlpvvlpdstlnlakpdfgaeakle 60

QY 61 VSSSCGPOCHKGTPLPYEAKOYLSVETLYANGSRTETOVGIYILSSSGDGAQHRDSDS 120

DB 61 vssscgpcqchkgtcplpyeakylsyetlyangstetqvgiylsssgdgaxhrdsdgs 120

QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNPFSTSVKLSGTCTGLTVAEKHVLTAACHIDHG 180

DB 121 sgksrrkrqiygydsrfsifgkdfllnypfstsvklsstgctglvaekhvltaachidhg 180

QY 181 KTYVKGTKLRVGFLLPKFKDGGRGANDSTAMPQKQFOWIRKTHVPGKWIKNAND 240

DB 181 ktyvkgtklrvgflpkfkdggrgandstampsqkqfowirvkrthvpgkwikgnand 240

QY 241 IGMDDYVALLLEKPKHKKRPMKIGVSPPAKQLPGGRHFHSGYDNDRPGNLVYRFCDVKDE 300

DB 241 igmddyvallelkpkhkrkfmkigvspapakqlpggrihfsgydnndrpgnlvyrfcdvkde 300

QY 301 TYDLLYQCCDAQPGASGSGVYVVRMWRKQOKWERKIIGIFSGHQWDMNGSPQDFNVAVR 360

DB 301 tydillyqcdagpgasggyvvyvmwrkrqqkwerkiigifsghqwvdmngspqdfnvavr 360

QY 361 ITPKVAQICYWIKGNVLDCREGDTVFPFGSN 392



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:15:34 ; Search time 35.78 Seconds  
(without alignments)  
240.882 Million cell updates/sec

Title: US-09-072-384-18  
Perfect score: 2080  
Sequence: 1 MAGIPGLLFLFLCAVQ.....LKVAICYWIKGNLYDCREG 383

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4	US-09-072-384-18
2	2080	100.0	392	4	US-09-072-384-15
3	2044	98.3	392	4	US-09-072-384-2
4	177	8.5	222	1	US-08-090-048-1
5	177	8.5	222	2	US-07-927-550-1
6	177	8.5	222	2	US-07-927-661A-1
7	122	5.9	256	3	US-08-906-769-89
8	122	5.9	256	3	US-08-906-616-89
9	122	5.9	256	4	US-08-817-795-89
10	122	5.9	256	4	US-08-639-075A-89
11	122	5.9	256	4	US-09-012-431-89
12	122	5.9	256	4	US-09-032-215-32
13	122	5.9	256	4	US-09-012-692-89
14	122	5.9	256	4	US-08-906-613-89
15	122	5.9	256	5	PCT-US95-1442A-89
16	117	5.6	437	1	US-08-487-037-2
17	115.5	5.6	241	4	US-08-944-483-59
18	114	5.5	228	4	US-08-944-483-44
19	114	5.5	253	6	5223425-8
20	113	5.4	238	6	5223425-5
21	113	5.4	250	6	5223425-4
22	112.5	5.4	223	1	US-08-278-091-13
23	112.5	5.4	223	1	US-08-483-859-13
24	112.5	5.4	223	1	US-08-472-173-13
25	112.5	5.4	223	2	US-08-487-167-13
26	112.5	5.4	223	2	US-08-482-816-13
27	112.5	5.4	223	2	US-08-296-149-13

28 112.5 5.4 223 2 US-08-801-499-13 Sequence 13, Appl  
29 112.5 5.4 223 3 US-08-615-271-13 Sequence 13, Appl  
30 112.5 5.4 223 3 US-09-074-660-13 Sequence 13, Appl  
31 112.5 5.4 223 3 US-09-074-659-13 Sequence 13, Appl  
32 112.5 5.4 223 3 US-09-106-468-13 Sequence 13, Appl  
33 112.5 5.4 223 4 US-09-106-468A-13 Sequence 13, Appl  
34 112.5 5.4 223 4 US-09-106-467-13 Sequence 13, Appl  
35 111.5 5.4 241 1 US-08-330-978-4 Sequence 4, Appl1  
36 111.5 5.4 241 1 US-08-474-042-4 Sequence 4, Appl1  
37 111.5 5.4 241 1 US-08-484-558-4 Sequence 4, Appl1  
38 111.5 5.4 241 1 US-08-774-592-4 Sequence 4, Appl1  
39 111.5 5.4 254 1 US-08-330-978-3 Sequence 3, Appl1  
40 111.5 5.4 254 1 US-08-474-042-3 Sequence 3, Appl1  
41 111.5 5.4 254 1 US-08-484-558-3 Sequence 3, Appl1  
42 111.5 5.4 254 1 US-08-774-592-3 Sequence 3, Appl1  
43 111.5 5.4 306 1 US-08-330-978-1 Sequence 1, Appl1  
44 111.5 5.4 306 1 US-08-474-042-1 Sequence 1, Appl1  
45 111.5 5.4 306 1 US-08-484-558-1 Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-072-384-18  
; Sequence 18, Application US/09072384  
; Patent No. 6153420  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES  
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,384  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, Gary E  
; REGISTRATION NUMBER: 31,648  
; REFERENCE/DOCKET NUMBER: 97-16C1  
; TELEPHONE: 206-442-6673  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 383 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Signal Sequence  
; LOCATION: 1..19  
; OTHER INFORMATION:  
US-09-072-384-18

```
Query Match 100.0%; Score 2080; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 5e-223;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFLCAVGVSPYSAPWKPTWPAYRLPVLPQSTLNLAKEPFGAEAKLE 60
Db 1 MAGIPGLLFLFLCAVGVSPYSAPWKPTWPAYRLPVLPQSTLNLAKEPFGAEAKLE 60
Qy 61 VSSCGPQCHKGTPPTVEEAKQYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
Db 61 VSSCGPQCHKGTPPTVEEAKQYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
Qy 121 SGKRRRQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEKHVLTAACHIDG 180
Db 121 SGKRRRQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEKHVLTAACHIDG 180
Qy 181 KTYVKGTKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRKTRHVPKGIKGNAND 240
Db 181 KTYVKGTKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRKTRHVPKGIKGNAND 240
Qy 241 IGMDDYALLELKKPKRKKFMKIGVSPPAKOLPGRIHFGYDNDPGLNLYRCDVKDE 300
Db 241 IGMDDYALLELKKPKRKKFMKIGVSPPAKOLPGRIHFGYDNDPGLNLYRCDVKDE 300
Qy 301 TYDLLYQCDQAQPGASGVVYRMKQKQKWKRIIGIFSGHQMVDNMGSPQDFNVAVR 360
Db 301 TYDLLYQCDQAQPGASGVVYRMKQKQKWKRIIGIFSGHQMVDNMGSPQDFNVAVR 360
Qy 361 ITPLKYAQCWIKGNVLDREG 383
Db 361 ITPLKYAQCWIKGNVLDREG 383

RESULT 2
US-09-072-384-15
; Sequence 15, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
```

```
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1...19
OTHER INFORMATION:
US-09-072-384-15

Query Match 100.0%; Score 2080; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 5.2e-223;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFLCAVGVSPYSAPWKPTWPAYRLPVLPQSTLNLAKEPFGAEAKLE 60
Db 1 MAGIPGLLFLFLCAVGVSPYSAPWKPTWPAYRLPVLPQSTLNLAKEPFGAEAKLE 60
Qy 61 VSSCGPQCHKGTPPTVEEAKQYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
Db 61 VSSCGPQCHKGTPPTVEEAKQYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
Qy 121 SGKRRRQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEKHVLTAACHIDG 180
Db 121 SGKRRRQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEKHVLTAACHIDG 180
Qy 181 KTYVKGTKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRKTRHVPKGIKGNAND 240
Db 181 KTYVKGTKLRVGLPKPKFDGGRGANDSTSAMPEQMKFQWIRKTRHVPKGIKGNAND 240
Qy 241 IGMDDYALLELKKPKRKKFMKIGVSPPAKOLPGRIHFGYDNDPGLNLYRCDVKDE 300
Db 241 IGMDDYALLELKKPKRKKFMKIGVSPPAKOLPGRIHFGYDNDPGLNLYRCDVKDE 300
Qy 301 TYDLLYQCDQAQPGASGVVYRMKQKQKWKRIIGIFSGHQMVDNMGSPQDFNVAVR 360
Db 301 TYDLLYQCDQAQPGASGVVYRMKQKQKWKRIIGIFSGHQMVDNMGSPQDFNVAVR 360
Qy 361 ITPLKYAQCWIKGNVLDREG 383
Db 361 ITPLKYAQCWIKGNVLDREG 383

RESULT 3
US-09-072-384-2
; Sequence 2, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:16:53 ; Search time 64.47 Seconds  
(without alignments)  
440.051 Million cell updates/sec

Title: US-09-072-384-18  
Perfect score: 2080  
Sequence: 1 MAGIPGLFLFLFLCAVCG.....LKQAICYWIKGNLYDCREG 383

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101:\*

1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT:\*

2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT:\*

3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT:\*

4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT:\*

5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT:\*

6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT:\*

7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT:\*

8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT:\*

9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT:\*

10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT:\*

11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT:\*

12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT:\*

13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT:\*

14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT:\*

15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT:\*

16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT:\*

17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT:\*

18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT:\*

19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT:\*

20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT:\*

21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT:\*

22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	20	Human transmembran
2	2080	100.0	383	20	WO927094 Seq ID 1
3	2080	100.0	383	20	Amino acid sequenc
4	2080	100.0	383	21	Protein encoded by
5	2080	100.0	383	21	Protein encoded by
6	2080	100.0	383	21	Human TANGO 186 pr
7	2080	100.0	383	21	Human signal pepti
8	2080	100.0	383	21	A bone marrow secr
9	2080	100.0	383	22	Human PRO307 prote
10	2080	100.0	383	22	Human Zsig13 varia
11	2080	100.0	392	22	Human Zsig13 varia

12	2073	99.7	383	22	Human membrane or
13	2044	98.3	392	22	Human Zsig13 varia
14	1909.5	91.8	375	21	Protein encoded by
15	1909.5	91.8	413	21	Protein encoded by
16	1890.5	90.9	382	21	Murine TANGO 186 p
17	1041	50.0	413	20	Human PRO1057 prot
18	1041	50.0	413	21	Human nn320_2 secr
19	1041	50.0	413	20	Human PRO1057 (UNQ
20	202.5	9.7	314	22	Bacillus lichenifo
21	194	9.3	316	13	Protease Blase. S
22	193	9.3	316	22	Bacillus lichenifo
23	184	8.8	222	22	Bacillus lichenifo
24	183	8.8	222	22	Bacillus lichenifo
25	183	8.8	222	22	Bacillus lichenifo
26	183	8.8	222	22	Bacillus lichenifo
27	183	8.8	222	22	Bacillus lichenifo
28	183	8.8	222	22	Bacillus lichenifo
29	181	8.7	222	22	Bacillus lichenifo
30	181	8.7	222	22	Bacillus lichenifo
31	181	8.7	222	22	Bacillus lichenifo
32	181	8.7	222	22	Bacillus lichenifo
33	180	8.7	222	22	Bacillus lichenifo
34	180	8.7	222	22	Bacillus lichenifo
35	180	8.7	222	22	Bacillus lichenifo
36	180	8.7	222	22	Bacillus lichenifo
37	179	8.6	222	22	Bacillus lichenifo
38	179	8.6	222	22	Bacillus lichenifo
39	179	8.6	222	22	Bacillus lichenifo
40	179	8.6	222	22	Bacillus lichenifo
41	179	8.6	222	22	Bacillus lichenifo
42	179	8.6	222	22	Bacillus lichenifo
43	179	8.6	222	22	Bacillus lichenifo
44	179	8.6	222	22	Bacillus lichenifo
45	179	8.6	222	22	Bacillus lichenifo

ALIGNMENTS

RESULT 1  
AA08657  
ID AA08657 standard; Protein; 383 AA.

AC AA08657;

DT 09-AUG-1999 (first entry)

DE Human transmembrane domain containing protein from clone HP10493.

KW Transmembrane domain; human; nutrition; cytokine; cell differentiation;

KW immune stimulation; immune suppression; haematopoiesis; activin;

KW regulatory tissue growth; inhibitor; chemostatic; chemokinetic;

KW haemostatic; thrombolytic; tumour inhibitor; anti-inflammatory;

KW gene therapy; screening.

OS Homo sapiens.

PN WO927094-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-JP05238.

PR 25-NOV-1997; 97JP-0323129.

PA (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENT.

PI Kato S, Kimura T, Sekine S;

DR WPI: 1999-357835/30.

DR N-PSDB; AAX77690, AAX77691.

XX



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:25:50 ; Search time 84.84 Seconds  
(without alignments)  
660.329 Million cell updates/sec

Title: US-09-072-384-18

Perfect score: 2080.  
Sequence: 1 MAGIPGLFLFLFLCAVQ.....LKAYQICYWIKGNLYDCREG 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

SPTREMBL17.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4	Q95084
2	1884.5	90.6	382	11	Q9D6X6
3	1042.5	50.1	413	4	Q9BQP6
4	133.5	6.4	799	11	Q9DRI0
5	129.5	6.2	303	2	Q9EXR9
6	129	6.2	469	6	Q9GMD9
7	127.5	6.1	1322	5	Q9NAT0
8	126.5	6.1	678	11	Q9JJS8
9	124.5	6.0	1322	5	Q9NJ55
10	124	6.0	266	6	Q46644
11	123.5	5.9	339	11	Q9QX91
12	123.5	5.9	366	11	Q9QX85
13	123.5	5.9	541	11	Q9QX90
14	123.5	5.9	623	11	Q9JJP3
15	123.5	5.9	643	11	Q9QX84
16	122	5.9	259	5	Q9XJG1
17	122	5.9	449	5	Q9VD08
18	120.5	5.8	482	11	Q63207
19	118	5.7	1376	5	Q9VQR8

20	117	5.6	258	5	Q9W508
21	117	5.6	522	5	Q01771
22	115.5	5.6	266	11	Q9D936
23	115.5	5.6	269	4	Q14243
24	115	5.5	258	4	Q9UN11
25	113.5	5.5	418	5	Q9VA87
26	113.5	5.5	685	11	Q9Z338
27	113.5	5.5	737	13	Q90422
28	113	5.4	573	5	Q9V516
29	112.5	5.4	603	5	Q9VDV1
30	112.5	5.4	686	4	Q9BZH0
31	112	5.4	249	13	Q9W7Q1
32	110.5	5.3	279	5	Q96991
33	110.5	5.3	481	11	Q54740
34	110	5.3	1449	5	Q10922
35	109.5	5.3	405	2	Q69973
36	108	5.2	274	5	Q17086
37	108	5.2	750	13	Q9W633
38	108	5.2	1047	5	Q24019
39	107.5	5.2	236	11	Q9Z1H1
40	107.5	5.2	427	10	Q9AYR4
41	107.5	5.2	481	11	Q88947
42	107.5	5.2	481	11	Q99L32
43	107.5	5.2	611	5	Q9VZS8
44	106	5.1	460	5	Q9VFZ6
45	106	5.1	762	13	Q9YIC6

#### ALIGNMENTS

RESULT 1  
O95084 PRELIMINARY; PRT; 383 AA.  
AC O95084;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SERINE PROTEASE (HYPOTHETICAL 43.0 KDA PROTEIN) (PROTEASE, SERINE,  
DE 23).  
GN ZSG13 OR DKFZP586B0719.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-UMBILICAL VEIN;  
RA Li X., Tedder T.F.;  
RT "A novel serine protease from human umbilical vein endothelial  
RT cells";  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-UTERUS;  
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE-CERVIX CARCINOMA;  
RA Strausberg R.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
DR EMBL; AF015287; AAD01553.1;  
DR EMBL; AF193611; AAF07186.1;  
DR EMBL; AL136914; CAB66848.1;  
DR EMBL; BC001278; AAH01278.1;  
DR MEROPS; S01.309; -.



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:24:07 ; Search time 32.13 seconds  
(without alignments)  
447.327 Million cell updates/sec

Title: US-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGLFLFLFLCAVQ.....IKGYLDCRGDTVPFPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	194	9.1	316	1	GSEP_BACLI	P80057	bacillus li
2	135.5	6.4	313	1	MPR_BACSU	P39790	bacillus su
3	124	5.8	286	1	EL1_BOVIN	Q28153	bos taurus
4	121.5	5.7	490	1	FA10_RABIT	O19045	oryctolagus
5	119	5.6	269	1	EL2_PIG	P08419	sus scrofa
6	115.5	5.4	269	1	EL2A_HUMAN	P08217	homo sapien
7	115.5	5.4	271	1	EL2_MOUSE	P05208	mus musculus
8	115	5.4	376	1	FA10_TROCA	P81428	tropidechis
9	114.5	5.4	238	1	TRY5_AEDAE	P29787	aedes aegyyp
10	113	5.3	266	1	EL1_PIG	P00772	sus scrofa
11	113	5.3	266	1	EL1_RAT	P00773	rattus norv
12	112.5	5.3	259	1	TRYP_STRGR	P00775	streptomyce
13	112.5	5.3	273	1	YDGD_ECOLI	P76176	escherichia
14	111.5	5.2	488	1	FA10_HUMAN	P00742	homo sapien
15	111.5	5.2	875	1	NETR_HUMAN	P56730	homo sapien
16	111	5.2	253	1	CFAD_HUMAN	P00746	homo sapien
17	111	5.2	761	1	NETR_MOUSE	O08762	mus musculus
18	110	5.2	269	1	EL2B_HUMAN	P08218	homo sapien
19	110	5.2	492	1	FA10_BOVIN	P00743	bos taurus
20	108	5.1	583	1	CFAL_HUMAN	P05156	homo sapien
21	108	5.1	786	1	STUB_DROME	Q05319	drosophila
22	107.5	5.0	686	1	MAS2_HUMAN	O00187	homo sapien
23	107	5.0	269	1	EL2_BOVIN	Q29461	bos taurus
24	107	5.0	603	1	CFAL_MOUSE	O61129	mus musculus
25	105.5	5.0	604	1	CFAL_RAT	O94443	rattus norv
26	105	4.9	274	1	TRY5_ANOGA	P35039	anopheles g
27	104.5	4.9	268	1	TRYP_STRGA	Q54179	streptomyce
28	103.5	4.9	271	1	EL2_RAT	P00774	rattus norv
29	103.5	4.9	416	1	FA9_BOVIN	P00741	bos taurus
30	103	4.8	855	1	ST14_HUMAN	Q94566	homo sapien
31	102	4.8	855	1	ST14_MOUSE	P56677	mus musculus
32	101.5	4.8	281	1	TRY2_DROER	P54630	drosophila
33	98.5	4.6	260	1	COGS_HYPLI	P08897	hypoderma l

## RESULT 1

GSEP\_BACLI  
ID GSEP\_BACLI STANDARD: PRT: 316 AA.  
AC P80057;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC ENDOPEPTIDASE) (GSE).  
GN BLASE.  
OS Bacillus licheniformis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1402;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=ATCC 14580;  
RX MEDLINE=93054737; PubMed=1429718;  
RA Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E., Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;  
RT "Purification, characterization, cloning, and expression of a glutamic acid-specific protease from Bacillus licheniformis ATCC 14580.";  
RT J. Biol. Chem. 267:23782-23788(1992).  
RL [2]  
RN SEQUENCE OF 95-316.  
RP MEDLINE=92155199; PubMed=1346764;  
RX Svendsen I., Bredam K.;  
RT "Isolation and amino acid sequence of a glutamic acid specific endopeptidase from Bacillus licheniformis.";  
RT Eur. J. Biochem. 204:165-171(1992).  
RL -!- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG PREFERENCE FOR GLU.  
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-XAA, GLU-XAA.  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL: D10060; BAA00949.1; -  
DR PIR: S23078; S23078.  
DR PIR: A45134; A45134.  
DR MEROPS: S01.271; -  
DR InterPro: IPR00126; Ser\_proteas\_v8.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00089; trypsin; 1  
DR PRINTS: PR00839; V8PROTEASE.  
DR SMART: SM00020; tryp\_Spc; 1.

## ALIGNMENTS

DR PROSITE; PS00672; V8\_HIS; 1.  
KW Hydrolyase; Serine protease; Signal.  
FT SIGNAL 1 ? 4  
FT PROPEP 1 ? 94  
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.  
FT ACT\_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 126 142  
FT DISULFID 275 279  
SQ SEQUENCE 316 AA; 96D7552CB7089B09 CRC64;

Query Match 9.1%; Score 194; DB 1; Length 316;  
Best Local Similarity 24.0%; Pred. No. 5e-09;  
Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;

QY 63 SSCGPOCHKTPL---PTYEAKQYLSYEITLYANGSRTEQVGIYILSSSGDGAHQHDSG 119  
DB 28 AQAPSPH--TPVSSDPY-KAETSVYDP-----NKSQYGLYSKAFGTGKVNTEKE 79

QY 120 SSGSRKRQIYGDYSFIFGKDFLEN-----YPFSTSVKLST---GCTGTLVAEKHV 170  
DB 80 KAERKSPAKAPY---STKSVIGSDRTRVNTTAYPYEIVHISSSIGSCTGWMIGPKTV 136

QY 171 LTAACHIDGKT-VYKQTKLQVGLPKPKDGRGANDSTSAMPEQMKQFWIRKRVTHV 229  
DB 137 ATAGHCITYDTSSGFACTATVSPG-----RNGTST-----YPYGSVKSTRFYI 178

QY 230 PKGWIKNANDIGMDYDYLLELKKPKHFKEMKIGVSPPAKLPGGRIHFSGYDNDPRGN 289  
DB 179 PSGRWNTN-----YDGAIELSEPIGNTVGVGYSTTSSLVGTVTTISGYPGDKTAG 233

QY 290 LVYRFCD--VKDETIDLYOCCDAQPCAGSGGVVVRMWRQOQKWKRII---GIFSGH 343  
DB 234 TQWQHSQPIAISEYIKLOYAM-DYVGQSGSPVFEQSSRTNCSGCSLAVHTNGVYGG- 291

QY 344 QWVDMNGSPQDNVAVRITPLKVAICYW 372  
DB 292 -----SSYNGRGTRETFEDNLTNW 311

RESULT 2  
MPR\_BACSU STANDARD; PRT; 313 AA.  
AC P39790;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE EXTRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).  
GN MPR.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=GP241;  
RX MEDLINE=90130256; PubMed=2105291;  
RA Sloma A., Rudolph C.F., Rufo G.A. Jr., Sullivan B.J., Theriault K.A., Ally D., Pero J.;  
RT "gene encoding a novel extracellular metalloprotease in Bacillus subtilis.";  
RL J. Bacteriol. 172:1024-1029(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;  
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the Bacillus subtilis chromosome.";  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-68 FROM N.A.

RX MEDLINE=89108019; PubMed=3145906;  
RA Smith H., de Jong A., Bron S., Venema G.;  
RT "Characterization of signal-sequence-coding regions selected from the Bacillus subtilis chromosome.";  
RL Gene 70:351-361(1988).  
RN [4]  
RP SEQUENCE OF 1-10 FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=961118702; PubMed=7496533;  
RA Saxild H.H., Jacobsen J.H., Nygaard P.;  
RT "Functional analysis of the Bacillus subtilis purT gene encoding formate-dependent glycinamide ribonucleotide transferase.";  
RL Microbiology 141:2211-2218(1995).  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.  
CC -1- CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2 FAMILY OF SERINE PROTEASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L10505; AAA22604.1; -;  
CC EMBL; AB006424; BAA33121.1; -;  
CC EMBL; M22916; AAA22832.1; -;  
CC EMBL; Z99105; CAB12018.1; -;  
CC PIR; A35122; A35122.  
CC HSP; P00756; 1SGF.  
CC MEROPS; S01.272; -;  
CC Subtilist; EG10690; mpr.  
CC InterPro; IPR000126; Ser\_proteas\_V8.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00839; V8PROTEASE.  
CC SMART; SM00020; Tryp\_SPC; 1.  
CC PROSITE; PS00672; V8\_HIS; 1.  
CC PROSITE; PS00673; V8\_SER; 1.  
KW Hydrolyase; Serine protease; Signal; Zymogen; Complete proteome.  
FT SIGNAL 1 34 POTENTIAL.  
FT PROPEP 35 93  
FT CHAIN 94 313 EXTRACELLULAR METALLOPROTEASE.  
FT ACT\_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 131 147 BY SIMILARITY.  
FT CONFLICT 61 68. QVSAPYEG -> PLESTAGA (IN REF. 3).  
SQ SEQUENCE 313 AA; 33842 MW; D41788B8D652AE94 CRC64;

Query Match 6.4%; Score 135.5; DB 1; Length 313;  
Best Local Similarity 22.2%; Pred. No. 0.00038;  
Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;

QY 72 GTPLPYEEAKQYLSYEITLYANGSRTEQVGIYILSSSGD--GAQHRDSSGSKSRKRQ 129  
DB 29 GVPAKAENQTSVNTGKREDAKNOT-----SKADQVSAPYEGTKTSKS----- 75

QY 130 IYGYDSRF-----SIFGKD-----FLNYPSTSVKLST-----GCTGTL 164  
DB 76 LYGGQTELEKNIQTLOPSSIICTDERTISSTTSPYRATVQLSIKYPNTSSYVGCTGL 135  
QY 165 VAEKHVLTAAHCIIH-----DGKTYVKGTKQLRVGLPKPKFKDGGRCAND 208  
DB 136 VNPNTVVTAGCHVSDHGWASTITAAPGRNGSSYPYGT----- 175  
QY 209 STSAMPEQMKQFWIRKRVHPKGIKGN---ANDIGMDYDYLLELKKPKHKKFMKIGV 265  
DB 176 -SGTMEYSVK-GWTESKDTNYDGAIKLNSPGNTVGVW-YGYRTNSSSP-----VGL 225



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:17:55 ; Search time 48.03 seconds  
(without alignments)  
621.704 Million cell updates/sec

Title: US-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGLFLFLFLCAVGO.....IKGNLYDCREGDTVFPPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5.

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.1	316	2 A45134	endopeptidase (EC
2	135.5	6.4	313	2 A35122	metalloproteinase
3	120.5	5.7	482	1 EXRT	coagulation factor
4	119	5.6	269	2 A26823	pancreatic elastase
5	117	5.5	522	2 T29767	hypothetical prote
6	115.5	5.4	269	2 B26823	pancreatic elastase
7	115.5	5.4	271	2 A25528	pancreatic elastase
8	115	5.4	258	2 S70439	pancreatic elastase
9	115	5.4	267	4 A56615	probable pancreati
10	114.5	5.4	238	1 TRWV5Y	trypsin-like prote
11	114	5.3	246	1 DBHU	complement factor
12	113	5.3	266	1 ELRT1	pancreatic elastase
13	113	5.3	266	1 ELPG	pancreatic elastase
14	112.5	5.3	259	1 TRSMG	trypsin (EC 3.4.21
15	112.5	5.3	273	2 E85765	hypothetical prote
16	112.5	5.3	273	2 H64915	hypothetical prote
17	111.5	5.2	488	1 EXHU	coagulation factor
18	111	5.2	761	2 JC5759	brain-specific ser
19	110	5.2	269	2 C26823	pancreatic elastase
20	110	5.2	492	1 EXBO	coagulation factor
21	110	5.2	1582	2 T15308	hypothetical prote
22	109.5	5.1	405	2 T35117	probable secreted
23	108	5.1	583	2 A29154	complement factor
24	108	5.1	786	1 A7547	serine proteinase
25	108	5.1	1047	2 A55617	masquerade precurs
26	107.5	5.0	236	2 A28566	T-cell suppressor
27	107.5	5.0	686	1 A59271	Ra-reactive factor
28	105	4.9	274	2 S40004	trypsin-related pr
29	103.5	4.9	271	1 ELRT2	pancreatic elastase

30	103.5	4.9	416	1 KEBO	coagulation factor
31	102.5	4.8	1238	2 T34929	hypothetical prote
32	100	4.7	272	2 JC4170	trypsin-like prote
33	98.5	4.6	267	2 S40006	trypsin (EC 3.4.21
34	98	4.6	548	2 D82175	probable trypsin V
35	98	4.6	624	2 T02289	probable polygalac
36	97.5	4.6	409	2 T35118	coagulation factor
37	97.5	4.6	452	1 A30351	complement factor
38	97.5	4.6	747	2 I51579	coagulation factor
39	97	4.6	1019	2 A38738	allergen Der f III
40	96.5	4.5	259	2 S68424	factor IX - rabbit
41	96.5	4.5	275	2 I46712	adenylate cyclase
42	96.5	4.5	2145	2 JC4747	serine proteinase
43	95.5	4.5	430	1 A24702	hypothetical prote
44	95	4.5	782	2 T32155	pancreatic elastase
45	94	4.4	268	2 S68825	pancreatic elastase

#### ALIGNMENTS

RESULT 1

A45134

endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis

C:Species: Bacillus licheniformis

C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999

C:Accession: A45134; S23078

R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S

J. Biol. Chem. 267, 23782-23786 (1992)

A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s

A:Reference number: A45134; MUID:93054737

A:Accession: A45134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <KAK>

A:CROSS-references: GB:D10060; NID:9216263; PIDN:BAA00949.1; PID:d1001415; PID:g21626

A:Experimental source: ATCC 14580

A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIP:118785)

R:Svendsen, I.; Brøddam, K.

Eur. J. Biochem. 204, 165-171, 1992

A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase

A:Reference number: S23078; MUID:92155199

A:Accession: S23078

A:Status: preliminary

A:Molecule type: protein

A:Residues: 95-316 <SVE>

C:Keywords: hydrolase

Query Match 9.1%; Score 194; DB 2; Length 316;  
Best Local Similarity 24.0%; Pred. No. 1.4e-08;  
Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;

QY 63 SSCGQCHKGTP---PYEEAKQVLSYETLYANGSRPTQVGIYILSSGGCAHQHDSG 119

DB 28 AQAAPSPH--TPVSSDPSY-KAETSVYDP-----NIKSDQYGLSKAFPTGKVNETKE 79

QY 120 SSGKRRKROIYGDERSIFCKDFLN-----YFSTSVKLST---GCTGLVAEKHV 170

DB 80 KAEKSPAKAPY---SIKSVIGSDDRTRVNTATPYRAIVHISSIGSCTGWMGPKTV 136

QY 171-LPAHETHDGT-YVKGTKLVGFLKPKFKDGRANDSTSAMPEQMFWIRVKRTHV 229

DB 137 ATAGHCIVDTSSGSPAGTATVSPG-----RNGTS-----YPGSVKSTRVFI 178

QY 230 PAKWTGNANDIMDYDVALLELKKPKFKMKIGVSPPAKOLPGRIHFSGYDNDPRGN 289

DB 179 PSGMRSGNTN-----YDGAIELSEPIGNTVGYFYGYSTTSLSLVGTVTITVSGPGDKTAG 233

QY 290 LVYRCD--TKDFTYLLVQQGDAQPGASGSGVYVRMNRQOQKWERKII-----GIFSCH 343

DB 234 TQWQHSGPIASETNKLQIAM-DTYGGSGSPVFPQSSRTNCSGPCSLAVHTNGVYGG- 291

QY 344 QWVDMNGSPQDFNVAVRITPLKYAICYW 372  
C:Species: Bacillus subtilis  
C:Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 15-Oct-1999  
C:Accession: A35122; I40010; A69660  
R:Sloma, A.; Rudolph, C.F.; Rufo Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; B  
J. Bacteriol. 172, 1024-1029, 1990  
A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.  
A:Reference number: A35122; MUID:90130256  
A:Accession: A35122  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-313 <SLO>  
A:CROSS-references: GB:110505; NID:gl43209; PIDN:AAA22604.1; PID:gl43210; GB:M29036  
R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.  
Gene 70, 351-361, 1988  
A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s  
A:Reference number: I39994; MUID:89108019  
A:Accession: I40010  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-60, 65, 'L', 67, 'S', 69, 'AQA' <RES>  
A:CROSS-references: GB:M22916; NID:gl43701; PIDN:AAA22832.1; PID:gl43702  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Buschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardinois,  
Y. M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Takakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69560  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-313 <KUN>  
A:CROSS-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12018.1; PID:el182176;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: mpr  
C:Keywords: hydrolase

Query Match 6.4%; Score 135.5; DB 2; Length 313;  
Best Local Similarity 22.2%; Pred.No. 0.001;  
Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;  
QY 72 GTPLPYEEAKQYLVETLYANGSRRTQVGIYLLSSGD--GAQHRDSSGSKSRKRQ 129  
Db 29 GVPKAAENPQTSVNSGKADATKNQT-----SKADQVAPYEGTGKTSK----- 75  
QY 130 IYGVDSRF-----STFGKD-----FLNYPSTSVKLS-----GCTGTL 164  
Db 76 LYGGQELEKNITLOPSSIIIGTDERTRISSTSFPRATVQUSIKYPTNYSYTGCTGFL 135  
QY 165 VAEKHVLTAAHCII-----DGKTYVKGQKLRVGLFKPKFDGGRGAND 208  
Db 136 VNPNTVVTAGHCYVSDHGWSHTTAAPGRNGSSYPYGV----- 175  
QY 209 STAMPEQMKFQIRVKRTHVPGKWIKN---ANDIGMDYDVALLELKPDKRKFMKIGV 265

Db 176 -SGTMFYSVR-GWTESKDTNYDYGAIKLNGSPGNTVGW-YGYRTTNSSP-----VGL 225  
QY 266 SPPAKQLPGRIHFSGYDNDPRGNLVYFCDVRDETDLLYQCCDAQPGASGSGVYRMW 325  
Db 226 SSVTGFPCDKTFTGTMWSDTKPIR-----SAETVKLTY-TTDTYGCOSGSPVY----- 272  
QY 326 KROQKWERRIIGIFSGHQMVMNGSPQDFNVAVRITPLKYAICYW 372  
Db 273 -RNYSDTGQTAIAHT-----NGG-SSYNLGTRTVNDVFNIIQYW 310  
RESULT 3  
EXRT  
coagulation factor Xa (EC 3.4.21.6) precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Jan-1995 #sequence\_revision 07-Feb-1997 #text\_change 08-Dec-2000  
C:Accession: S49075; JC4670; PS0191; PS0190; I62745  
R:Stanton, C.; Ross, P.; Hutson, S.; Wallin, R.  
Thromb. Res. 80, 63-73, 1995  
A:Title: Evidence for competition between vitamin K-dependent clotting factors in  
A:Reference number: A58498; MUID:96093366  
A:Accession: S49075  
A:Molecule type: mRNA  
A:Residues: 1-482 <STAI>  
A:CROSS-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
A:Note: submitted to the EMBL Data Library, June 1994  
A:Note: neither the complete nucleic acid sequence nor the complete translation are s  
R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.  
Gene 169, 269-273, 1996  
A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.  
A:Reference number: JC4670; MUID:96194815  
A:Accession: JC4670  
A:Molecule type: mRNA  
A:Residues: 1-482 <STAI>  
A:CROSS-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
A:Experimental source: Cos-1 cell  
R:Enyoji, K.; Miyazaki, K.; Kato, H.  
J. Biochem. 109, 890-898, 1991  
A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat  
A:Reference number: PS0190; MUID:92041742  
A:Accession: PS0191  
A:Molecule type: protein  
A:Residues: 41-58, 'X', 60-65 <ENJ1>  
A:Accession: PS0190  
A:Molecule type: protein  
A:Residues: 183-186, 'X', 188-207 <ENJ2>  
R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.  
Eur. J. Haematol. 52, 162-168, 1994  
A:Title: Analysis of the partial nucleotide sequences and deduced primary struc  
A:Reference number: I46196; MUID:94222160  
A:Accession: I62745  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 295-383, 'G', 385-455 <MUR>  
A:CROSS-references: GB:D21215; NID:g415309; PIDN:BAA04756.1; PID:g455396  
C:Function:  
A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the  
A:Pathway: blood coagulation  
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol  
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-40/Domain: propeptide #status predicted <PRO>  
F:25-84/Domain: Gla domain homology <GLA>  
F:41-179/Product: coagulation factor X light chain #status predicted <LCH>  
F:90-121/Domain: EGF homology <EG1>  
F:129-164/Domain: EGF homology <EG2>  
F:183-482/Product: coagulation peptide #status predicted <HCH>  
F:183-231/Domain: activation peptide #status predicted <APT>  
F:232-482/Product: coagulation factor Xa heavy chain #status predicted <ACT>  
F:232-460/Domain: trypsin homology <TRY>  
F:46,47,54,56,59,60,65,66,69,72,79/Modified site: gamma-carboxyglutamic acid (Glu) #s  
F:57-62,90-101,95-110,112-121,129-140,136-149,151-164,172-340,238-243,259-275,388-402

C; Accession: T29767  
R; Waterston, B.; Gattung, S.; Le, T.T.

C; Accession: T29767  
R; Waterston, B.; Gattung, S.; Le, T.T.

A;Molecule type: mRNA  
A;Residues: 1-269 <FLE>

A;Molecule type: mRNA  
A;Residues: 1-269 <FLE>

Qy	148	YPFSTSVK-LSTG-----CTGTLVAGHVLTAAC	HDKGTY--VKG	TOKLRVGF	LKPKF	199
Db	42	WPOVSLQVLSGRWRHNCSSGVANNWLTAAHCL	ENTQTV	RVLLGAHSL		92
Qy	200	KDGGRGANDSTSAMPEQMFWLRVRKTRVPK	GWIKGNANDIGMDYD	YAL	LELKKP-HKR	258
Db	93	--SNPGAGSAA-----VQSEKLVVHQW----	NSQNVGNGVDIAL	IKLASPTVLS		136
Qy	259	KFMKIGVSPPAKO-LPGRI-HFSGY-----	DNDRPGNLV----	YRFC-----		295
Db	137	KNIQTACLPAGTILPRNYCYVYTGWGLLTQ	NGNSPDLRQGRLLVVD	YATCSSASWGS		196
Qy	296	DVADETY---DLLYQOCDAQCSGCVYVRMK	QKQKWERKIKIFSG	HQWVDMGS		351
Db	197	SVKSSMWACGGDGVTSSEN---GDSGGPLNCR	ASNGQ---WQ---VHGIVSFGSSLCGN-Y			247
Qy	352	PODFNVAVRITPLKYAICYWIKGNYLD		379		
Db	248	PKRPSVETRVS-----NVID		262		

RESULT 8  
S70439  
pancreatic elastase I (allele HEL1-16) probable splice form I - human  
C; Species: Homo sapiens (man)  
C; Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 16-Dec-1998  
C; Accession: S70439

A: Accession: S70439  
A: Status: preliminary; not compared with conceptual translation  
A: Molecule type: DNA  
A: Residues: 1-258 <RAW>  
C: Superfamily: trypsin; trypsin homology  
F: 19-251/Domain: trypsin homology <TRY>

```

QY 199 FKDGGRGANDSTAMPQMKFWIRVKRTHVPKGWTKGNANDIGMDYDVALLLEL-KKPKHK 257
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 81 -----SQNDGTE-----QYVSOKIVVHPYW---NSDNVAAGYDIALRLAQSVTL 123
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 258 RKFMKIGKGVSP 267
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 124 NSYVQLGVLP 133
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
A$6615
probable pancreatic elastase (EC 3.4.21.36) pseudogene - human
N;Alternate names: pancreatic elastase I homolog; pancreatic elastase I al
C;Species: Homo sapiens (man)
C;Date: 11-Aug-1995 #sequence_revision 17-Aug-1995 #text_change 14-Aug-1995
C;Accession: A56615; S70440
R;Kawashima, I.; Tani, T.; Mita-Honjo, K.; Shimoda-Takano, K.; Ohmine, T.;
DNA Seq. 2, 303-312, 1992
A;Title: Genomic organization of the human homologue of the rat pancreatic
A;Reference number: A56615; MUID:92338395

```

A; Molecule type: Dna  
A; Residues: 1-267 <RAW>  
A; Cross-references: EMBL:X62259; NID:g31246; EMBL:X62258; GB:S40923; NID:g31247; EMBL:  
5; GB:S40856; NID:g31251; EMBL:X62256; GB:S40857; NID:g31252; EMBL:X62257; GB:S40859;  
A; Note: sequence extracted from NCBI backbone (NCBIN:109315, NCBIN:109317, NCBIN:109318)

C:Comment: This apparently silent human homolog of pancreatic elastase I is a single-copy functional protein in some other tissue.

C:Genetics:  
A:Gene: GDB:ELAI  
A:Map position: 12  
C:Keywords: hydrolase; pseudogene; serine proteinase

Query Match 5.4%; Score 115; DB 4; Length 267;  
Best Local Similarity 26.9%; Pred. No. 0.042;  
Matches 35; Conservative 23; Mismatches 38; Indels 34; Gaps 6;

QY 147 NYPFSTSVKLSGT-----CTGTLVAEKHVLTAACHIDGKYY--VKGQKLRVGLPKPK 198

DB 38 SWPSQISLQYRSGSWYHTCGGTLIRQNNWMTAAHCVDYQKTRFVAGDHNL----- 89

QY 199 FKDDGGGANDSTSAMPEQMKFQIRVKRTHVPKGIKGNANDIGMDYDVALLEL-KKPKH 257

DB 90 -----SQNDGTE-----QYVSQKIVVHPYW---NSDINVAAGYDIALRLAQSVTL 132

QY 258 RKFMKIGVSP 267

DB 133 NSYVQLGVLP 142

#### RESULT 10

TRWV3Y

Trypsin-like proteinase (EC 3.4.21.-) 5G1 precursor - yellow fever mosquito (fragment)

C:Species: Aedes aegypti (yellow fever mosquito)

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 18-Jun-1999

C:Accession: S19891

R:Kahnok, S.; Tabak, L.M.; Prosser, D.E.; Downe, A.E.R.; White, B.N.

submitted to the EMBL Data Library, February 1992

A:Description: Isolation, sequencing and characterization of 2 cDNA clones coding for tr

A:Reference number: S19890

A:Accession: S19891

A:Molecule type: mRNA

A:Residues: 1-238 <KAL>

A:Cross-references: EMBL:X64363; NID:g5563; PIDN:CAA45715.1; PID:g5564

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; insect midgut; protein digestion; serine proteinase

F:1-11/Domain: signal sequence and propeptide (fragment) #status predicted <SIG>

F:12-238/Product: trypsin-like proteinase 5G1 #status predicted <MAT>

F:12-232/Domain: trypsin homology <TRY>

F:38-54/162-178,189-213/Disulfide bonds: #status predicted

F:53.97,193/Active site: His, Asp, Ser #status predicted

Query Match 5.4%; Score 114.5; DB 1; Length 238;  
Best Local Similarity 29.4%; Pred. No. 0.04;  
Matches 42; Conservative 22; Mismatches 44; Indels 35; Gaps 7;

QY 149 PFSTSTSVKLSGT---VKLSTGCTGTLVAEKHVLTAACHIDGKTYVKGTQKLRVGLPKPKRKG 205

DB 24 PQVSLSGVSGSHFCGSLLSRWYWTAGCAASGQTNL----QVRIG--SSQHASGG-- 75

QY 206 ANDSTSAMPEQMKFQIRVKRTH-VPKGIKGNANDIGMDYDVALLELKKPKRKFMTG 264

DB 76 -----OLIKVKVNRHPK-----YDEVTTDYDFALLELEETVTFSDSCAP 115

QY 265 VSPPAKQLP---GGRHFHFGYDN 284

DB 116 VKLPQKDTVPNEGTCLOVSGWGN 138

#### RESULT 11

DBHU

Complement factor D (EC 3.4.21.46) precursor [validated] - human (fragment)

N:Alternate names: adipsin; C3 convertase activator

C:Species: Homo sapiens (man)

C:Date: 28-Aug-1985 #sequence\_revision 31-Dec-1992 #text\_change 08-Dec-2000

C:Accession: A40197; A00936; A60571; S66645

R:White, R.T.; Damm, D.; Hancock, N.; Rosen, B.S.; Lowell, B.B.; Usher, P.; Flier, J.S.;

J. Biol. Chem. 267, 9210-9213, 1992

A:Title: Human adipsin is identical to complement factor D and is expressed at high l

A:Reference number: A40197; MUID:92250520

A:Accession: A40197

A:Molecule type: mRNA

A:Residues: 1-246 <WHI>

A:Cross-references: GB:M84526

R:Niemann, M.A.; Bhowm, A.S.; Bennett, J.C.; Volanakis, J.E.

Biochemistry 23, 2482-2486, 1984

A:Title: Amino acid sequence of human D of the alternative complement pathway.

A:Reference number: A00936; MUID:85000441

A:Accession: A00936

A:Molecule type: protein

A:Residues: 19-44,'G',46-51,'Q',53-75,'TH',78,'P',80-83,'XXXITIE',90-172,86-91,185-23

A>Note: a few residues were assigned from the previously published sequence of Reid e

R:Miya, T.; Oda, O.; Inagi, R.; Sugiyama, S.; Miyama, A.; Maeda, K.; Nakashima, I.;

Mol. Immunol. 27, 637-644, 1990

A:Title: Molecular and functional identification and purification of complement compo

A:Reference number: A60571; MUID:90370044

A:Accession: A60571

A:Molecule type: protein

A:Residues: 19-20,'XX',23-27,'XX',30-31,'XX',34,'X',36-40 <MIY>

R:Balk, N.; Holtkamp, U.; Hoerl, W.H.; Tschesche, H.

FEBS Lett. 371, 300-302, 1995

A:Title: Inhibition of degradation of human polymorphonuclear leukocytes by complem

A:Reference number: S66645; MUID:96013156

A:Accession: S66645

A:Molecule type: preliminary

A:Status: preliminary

A:Residues: 19-44,'C',46-48 <BAL>

C:Comment: Factor D cleaves factor B when the latter is complexed with factor C3b, ac

C:Genetics:

A:Gene: GDB:DF

A:Cross-references: GDB:132645; OMIM:134350

A:Map position: Xpter:Xqter

C:Superfamily: trypsin; trypsin homology

C:Keywords: complement alternate pathway; hydrolase; plasma; serine proteinase

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-246/Product: complement factor D (fragment) #status experimental <MAT>

F:19-241/Domain: trypsin homology <TRY>

F:44-60,141-207,172-188,197-222/Disulfide bonds: #status predicted

F:59,105,201/Active site: His, Asp, Ser #status predicted

Query Match 5.3%; Score 114; DB 1; Length 246;  
Best Local Similarity 28.4%; Pred. No. 0.046;  
Matches 44; Conservative 20; Mismatches 43; Indels 48; Gaps 9;

QY 149 PFSTSVKLSGT---CTGTLVAEKHVLTAACHIH---DGKTYVKGTKLRVGLPKPKFDGG 203

DB 31 PYMASVOLNGAHLCAGLVAERWVLSAAHCLDADADGKVQVL----- 72

QY 204 RGANDSTSAMPEQMKFQIRVKRTHVPKGIKGNANDIGMDYDVALLELKKPKRKFMTK 263

DB 73 LGAHLSLQPEPSKRLYDVURA---VP-----HPDSQPTDIDHLLLLQLSE-----KA 117

QY 264 GVSPPAKQLPGRHIFSGYDND--RPGNLVYRFCDV 297

DB 118 TLGPAVRPLPWQVR-----DROVAPQTL---CDV 143

#### RESULT 12

ELRT1

pancreatic elastase (EC 3.4.21.36) I precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 18-Aug-1982 #sequence\_revision 18-Aug-1982 #text\_change 18-Jun-1999

C:Accession: A00960; A20534

R:MacDonald, R.J.; Swift, G.H.; Quinto, C.; Swain, W.; Pictet, R.L.; Nikovits, W.;

Biochemistry 21, 1453-1463, 1982

A:Title: Primary structure of two distinct rat pancreatic preproelastases determined

A:Reference number: A00960; MUID:82182967

A:Accession: A00960

A:Molecule type: mRNA

A:Accession: A20777  
A:Molecule type: mRNA  
A:Residues: 1-125, 'G', 127-183, 'L', 185-266 <TAN>  
A:Cross-references: GB:D00160; NID:G217683; PIDN:BAA00118.1; PID:G217684  
A:Note: the authors translated the codon GGC for residue 58 as Gln, GGC for residue 126  
R:Shotton, D.M.; Hartley, B.S.  
Biochem. J. 131, 643-675, 1973  
A:Title: Evidence for the amino acid sequence of porcine pancreatic elastase.  
A:Reference number: A90267; MUID:7329121  
A:Accession: A10061  
A:Molecule type: protein  
A:Residues: 27-91, 'N', 93-203, 'N', 205-266 <SHO>  
R:Shotton, D.M.; Hartley, B.S.  
Nature 225, 811-816, 1970  
A:Title: Three-dimensional structure of tosyl-elastase.  
A:Reference number: A93160; MUID:70114044  
A:Contents: annotation; x-ray crystallography, 3.5 angstroms; active site  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; pancreas; serine proteinase; zymogen  
F:1-16/Domain: signal sequence status predicted &ltSIG>

A:Reference number: A93160; MUID:70114044  
A:Contents: annotation; X-ray crystallography, 3.5 angstroms; active site  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; pancreas; serine proteinase  
F.1.16/Domain: signal sequence #status predicted<SIG>

OV 208 DS--TSAMPEOMKFOWIRVKRTHVPKGVKGNANDIGMDYDYALLELKKPHKRKEMKIG 264

Db 82 TSTATGGVVDLQSSAVKVRSTKVLQA-----PGYNGTGKDWALIKLAQPINQPTLKIA 136  
QY 265 VSPPAKQLPGGRIFHSFYDNDPGLNLYRRF 294  
Db 137 TTTAYNQ---GTFYVAGWGANREGGSGQRY 163

RESULT 15  
E85765  
hypothetical protein 22592 [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: E85765  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: AB5480; MUID:21074935; PMID:11206551  
A:Accession: E85765  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-273 <STO>  
A:Cross-references: GB:AE005174; NID:g12515576; PIDN:AG56385.1; GSPDB:GN00145; UMGF:225  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: 22592

Query Match 5.3%; Score 112.5; DB 2; Length 273;  
Best Local Similarity 23.3%; Pred. No. 0.069;  
Matches 62; Conservative 35; Mismatches 112; Indels 57; Gaps 13;

QY 101 VGIYILSSGSGAQHRDSSGSKSRKROIYGYDSRFSTFGKDFLLNYPFSTSVKLSTG- 159  
Db 9 LCAISLTSAFVADKPDVAKSANDEVSTLFFGHDDRPV---NDTQSPWDVAVGQLETAS 65

QY 160 ---CTGTLVAEKHVLTAHCIHDGKTYVKGTKLRYGFLKPKFKDGR-----GANDS 209  
Db 66 GNLCTATLIAPNLALTAGHCL---LTPPKADKAVALRFVSNKGLWRYDIHDIEGRVDP 122

QY 210 TSAMPEOMKFWIRVKTHTVHPKWKGNANDIGMDYDIALLELKKPKRKFMKIGVSP-- 267  
Db 123 T--LGKRLRAD-----GDGWIVPPA---AAPWDFGLIVLRNPPS-----GITPLP 162

QY 268 -----PAKQLPGGRIFHSFYDNDPGLNLY--YRFCDVKD-ETYDLLYQCCDAQPGA 315  
Db 163 LFEQDKAALTAALKAGRKVTQAGYPEDHLDLTLYSHQNCVETGWAQTSVASHQCDTLPGD 222

QY 316 SSGSVYVRMWRKQKQKWKRIIGIFS 341  
Db 223 SGSPLML---HTDDGQWQ--LIGVQS 242

Search completed: December 2, 2001, 16:17:56  
Job time: 191 sec

***This Page Blank (uspto)***



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:14:40 ; Search time 35.78 Seconds  
(without alignments)  
246.542 Million cell updates/sec

Title: US-09-072-384-2  
Perfect score: 2112  
Sequence: 1 MAGIPGLFLFLLCVAVGQ.....IKGNYLDCREGDTVFPLPGSN 392

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2104	99.6	392	4	US-09-072-384-2
2	2085	98.7	392	4	US-09-072-384-15
3	2044	96.8	383	4	US-08-023-384-18
4	185	7.8	222	1	US-08-090-048-1
5	165	7.8	222	2	US-08-292-550-1
6	165	7.8	222	2	US-07-927-661A-1
7	123	5.8	256	3	US-08-906-769-89
8	123	5.8	256	3	US-08-906-616-89
9	123	5.8	256	4	US-08-817-795-89
10	123	5.8	256	4	US-08-639-075A-89
11	123	5.8	256	4	US-09-012-431-89
12	123	5.8	256	4	US-09-032-215-32
13	123	5.8	256	4	US-09-012-692-89
14	123	5.8	256	4	US-08-906-613-89
15	123	5.8	256	5	PCT-US95-1442A-89
16	121.5	5.8	437	1	US-08-487-037-2
17	118	5.6	241	4	US-08-944-483-59
18	116.5	5.5	437	1	US-08-487-037-3
19	114	5.4	254	1	US-08-330-978-3
20	114	5.4	254	1	US-08-474-042-3
21	114	5.4	254	1	US-08-484-558-3
22	114	5.4	254	1	US-08-774-592-3
23	114	5.4	306	1	US-08-330-978-1
24	114	5.4	306	1	US-08-474-042-1
25	114	5.4	306	1	US-08-484-558-1
26	114	5.4	306	1	US-08-774-592-1
27	114	5.4	448	1	US-08-295-411-3

28	114	5.4	448	2	US-08-955-471-3
29	114	5.4	448	5	PCT-US92-10068-1
30	114	5.4	448	5	PCT-US92-10242-3
31	114	5.4	487	1	US-08-469-486-53
32	114	5.4	487	2	US-08-469-658-53
33	114	5.4	486	1	US-08-487-037-1
34	114	5.4	492	2	US-08-469-486-2
35	114	5.4	492	2	US-08-469-658-2
36	113.5	5.4	241	1	US-08-330-978-4
37	113.5	5.4	241	1	US-08-474-042-4
38	113.5	5.4	241	1	US-08-484-558-4
39	113.5	5.4	241	1	US-08-774-592-4
40	112.5	5.3	223	1	US-08-278-091-13
41	112.5	5.3	223	1	US-08-483-859-13
42	112.5	5.3	223	1	US-08-472-173-13
43	112.5	5.3	223	2	US-08-487-167-13
44	112.5	5.3	223	2	US-08-482-816-13
45	112.5	5.3	223	2	US-08-296-149-13

ALIGNMENTS

RESULT 1  
US-09-072-384-2  
; Sequence 2, Application US/09072384  
; Patent No. 6153420  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES  
; NUMBER OF INVENTIONS: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,384  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, Gary E  
; REGISTRATION NUMBER: 31,648  
; REFERENCE/DOCKET NUMBER: 97-16C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6673  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 392 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Signal Sequence  
; LOCATION: 1..19  
; OTHER INFORMATION:  
US-09-072-384-2

```
Query Match 99.6%; Score 2104; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 6.6e-219;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFFLLCAGGVSPYSAPWKTWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
DB 1 MAGIPGLLFLFFLLCAGGVSPYSAPWKTWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
DB 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
DB 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
QY 181 KTVVGTQKLRVGLFKPKFKDGGGRANDSTSAMPEOMKQFOWIRKTHVPGWIKGNAND 240
DB 181 KTVVGTQKLRVGLFKPKFKDGGGRANDSTSAMPEOMKQFOWIRKTHVPGWIKGNAND 240
QY 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDRPGNLVYRFGCDVKDE 300
DB 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDRPGNLVYRFGCDVKDE 300
QY 301 TYDLLYQQCDAQPCASGYGVYVWMKRQOQKWERKIIGIFSGHGWMDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPCASGYGVYVWMKRQOQKWERKIIGIFSGHGWMDMNGSPQDFNVAVR 360
QY 361 ITPKLYAQICYWIKGNVLDRCRGDTVFLPGSN 392
DB 361 ITPKLYAQICYWIKGNVLDRCRGDTVFLPGSN 392

RESULT 2
US-09-072-384-15
; Sequence 15, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1...19
; OTHER INFORMATION:
US-09-072-384-15

Query Match 98.7%; Score 2085; DB 4; Length 392;
Best Local Similarity 98.2%; Pred. No. 7.5e-217;
Matches 395; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFFLLCAGGVSPYSAPWKTWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
DB 1 MAGIPGLLFLFFLLCAGGVSPYSAPWKTWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
DB 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSSGDGAAXRDSGS 120
QY 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
DB 121 SGKSRKRQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEXHVLTAACHIDG 180
QY 181 KTVVGTQKLRVGLFKPKFKDGGGRANDSTSAMPEOMKQFOWIRKTHVPGWIKGNAND 240
DB 181 KTVVGTQKLRVGLFKPKFKDGGGRANDSTSAMPEOMKQFOWIRKTHVPGWIKGNAND 240
QY 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDRPGNLVYRFGCDVKDE 300
DB 241 IGMDDYALLELKKPKHKKFKMKGVSPPAKQLPGGRHFGSYDNDRPGNLVYRFGCDVKDE 300
QY 301 TYDLLYQQCDAQPCASGYGVYVWMKRQOQKWERKIIGIFSGHGWMDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPCASGYGVYVWMKRQOQKWERKIIGIFSGHGWMDMNGSPQDFNVAVR 360
QY 361 ITPKLYAQICYWIKGNVLDRCRGDTVFLPGSN 392
DB 361 ITPKLYAQICYWIKGNVLDRCRGDTVFLPGSN 392

RESULT 3
US-09-072-384-18
; Sequence 18, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:15:33 ; Search time 35.78 seconds  
(without alignments)  
246.542 Million cell updates/sec

Title: US-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGLFLFLFLCAVQ.....IKGNVLDRCRGDTVPFPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2131	100.0	392	4	US-09-072-384-15
2	2085	97.8	392	4	US-09-072-384-2
3	2080	97.6	383	4	US-09-072-384-18
4	177	8.3	222	1	US-08-090-048-1
5	177	8.3	222	2	US-08-252-550-1
6	177	8.3	222	2	US-07-927-661A-1
7	122	5.7	256	3	US-08-906-769-89
8	122	5.7	256	3	US-08-906-616-89
9	122	5.7	256	3	US-08-817-795-89
10	122	5.7	256	4	US-08-639-075A-89
11	122	5.7	256	4	US-09-012-431-89
12	122	5.7	256	4	US-09-032-215-32
13	122	5.7	256	4	US-09-012-692-89
14	122	5.7	256	4	US-08-906-613-89
15	122	5.7	256	5	PCT-US95-14442A-89
16	117	5.5	437	1	US-08-487-037-2
17	115.5	5.4	241	4	US-08-944-483-59
18	114	5.3	228	4	US-08-944-483-44
19	114	5.3	253	6	5223425-8
20	113	5.3	238	6	5223425-5
21	113	5.3	250	6	5223425-4
22	112.5	5.3	223	1	US-08-278-091-13
23	112.5	5.3	223	1	US-08-483-859-13
24	112.5	5.3	223	1	US-08-472-173-13
25	112.5	5.3	223	2	US-08-487-167-13
26	112.5	5.3	223	2	US-08-482-816-13
27	112.5	5.3	223	2	US-08-296-149-13

28 112.5 5.3 223 2 US-08-801-499-13 Sequence 13, Appl  
29 112.5 5.3 223 2 US-08-615-271-13 Sequence 13, Appl  
30 112.5 5.3 223 3 US-09-074-660-13 Sequence 13, Appl  
31 112.5 5.3 223 3 US-09-074-659-13 Sequence 13, Appl  
32 112.5 5.3 223 3 US-09-106-468-13 Sequence 13, Appl  
33 112.5 5.3 223 4 US-09-106-466A-13 Sequence 13, Appl  
34 112.5 5.3 223 4 US-09-106-467-13 Sequence 13, Appl  
35 111.5 5.2 241 1 US-08-330-978-4 Sequence 4, Appl  
36 111.5 5.2 241 1 US-08-474-042-4 Sequence 4, Appl  
37 111.5 5.2 241 1 US-08-484-558-4 Sequence 4, Appl  
38 111.5 5.2 241 1 US-08-774-592-4 Sequence 4, Appl  
39 111.5 5.2 254 1 US-08-330-978-3 Sequence 3, Appl  
40 111.5 5.2 254 1 US-08-474-042-3 Sequence 3, Appl  
41 111.5 5.2 254 1 US-08-484-558-3 Sequence 3, Appl  
42 111.5 5.2 254 1 US-08-774-592-3 Sequence 3, Appl  
43 111.5 5.2 306 1 US-08-330-978-1 Sequence 1, Appl  
44 111.5 5.2 306 1 US-08-474-042-1 Sequence 1, Appl  
45 111.5 5.2 306 1 US-08-484-558-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-072-384-15  
; Sequence 15, Application US/09072384  
; Patent No. 6153420  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES  
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,384  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, Gary E  
; REGISTRATION NUMBER: 31,648  
; REFERENCE/DOCKET NUMBER: 97-16C1  
; TELEPHONE: 206-442-6673  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 392 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Signal Sequence  
; LOCATION: 1..19  
; OTHER INFORMATION:  
US-09-072-384-15

```
Query Match 100.0%; Score 2131; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.1e-227;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFLLCAGGVSPYSAPWKPWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
DB 1 MAGIPGLLFLFLLCAGGVSPYSAPWKPWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
QY 61 VSSSCGPQCHKGTPLYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
DB 61 VSSSCGPQCHKGTPLYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
QY 121 SGKSRRKQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEKHVLTAACHIDG 180
DB 121 SGKSRRKQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEKHVLTAACHIDG 180
QY 181 KTVVKTQKLRVGLFKPKFQKGGANDSTSAMPEQMKFQWIRKTHVPKGIKNAND 240
DB 181 KTVVKTQKLRVGLFKPKFQKGGANDSTSAMPEQMKFQWIRKTHVPKGIKNAND 240
QY 241 IGMDDYALLELKKPHKRFKMGVSPPAKQLPGRHFGSYDNDRPNLVYRFGDVKDE 300
DB 241 IGMDDYALLELKKPHKRFKMGVSPPAKQLPGRHFGSYDNDRPNLVYRFGDVKDE 300
QY 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHGWDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHGWDMNGSPQDFNVAVR 360
QY 361 ITPKLYAQICYWIKGNLYDCREGDTVFPFGSN 392
DB 361 ITPKLYAQICYWIKGNLYDCREGDTVFPFGSN 392
```

RESULT 2

```
US-09-072-384-2
; Sequence 2, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: zymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-16C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
```

```
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1...19
OTHER INFORMATION:
US-09-072-384-2

Query Match 97.8%; Score 2085; DB 4; Length 392;
Best Local Similarity 98.2%; Pred. No. 1.5e-222;
Matches 395; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFLLCAGGVSPYSAPWKPWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
DB 1 MAGIPGLLFLFLLCAGGVSPYSAPWKPWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60
QY 61 VSSSCGPQCHKGTPLYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
DB 61 VSSSCGPQCHKGTPLYEAKOYLSYETLYANGSRTEQVGIYILSSSGDGAQHRDGS 120
QY 121 SGKSRRKQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEKHVLTAACHIDG 180
DB 121 SGKSRRKQIYGYDSRFSIFGKDFLLNYPSTSVKLSGTGTLVAEKHVLTAACHIDG 180
QY 181 KTVVKTQKLRVGLFKPKFQKGGANDSTSAMPEQMKFQWIRKTHVPKGIKNAND 240
DB 181 KTVVKTQKLRVGLFKPKFQKGGANDSTSAMPEQMKFQWIRKTHVPKGIKNAND 240
QY 241 IGMDDYALLELKKPHKRFKMGVSPPAKQLPGRHFGSYDNDRPNLVYRFGDVKDE 300
DB 241 IGMDDYALLELKKPHKRFKMGVSPPAKQLPGRHFGSYDNDRPNLVYRFGDVKDE 300
QY 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHGWDMNGSPQDFNVAVR 360
DB 301 TYDLLYQQCDAQPGASGVYVVRMKRQKQKWERKIIGIFSGHGWDMNGSPQDFNVAVR 360
QY 361 ITPKLYAQICYWIKGNLYDCREGDTVFPFGSN 392
DB 361 ITPKLYAQICYWIKGNLYDCREGDTVFPFGSN 392

RESULT 3
US-09-072-384-18
; Sequence 18, Application US/09072384
; Patent No. 6153420
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: zymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,384
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:14:40 ; Search time 64.47 Seconds  
(without alignments)  
450.391 Million cell updates/sec

Title: US-09-072-384-2  
Perfect score: 2112  
Sequence: 1 MAGIPGLLFLFLCAVGO.....IKGNYLDCRGDVFPLPGSN 392

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	2104	99.6	392	22	Human Zsig13 varia
2	2085	98.7	392	22	Human Zsig13 varia
3	2044	96.8	383	20	Human transmembran
4	2044	96.8	383	20	WO9927094 Seq ID 1
5	2044	96.8	383	20	Amino acid sequenc
6	2044	96.8	383	21	Protein encoded by
7	2044	96.8	383	21	Protein encoded by
8	2044	96.8	383	21	Human TANGO 186 pr
9	2044	96.8	383	21	Human signal pepti
10	2044	96.8	383	21	A bone marrow secr
11	2044	96.8	383	22	Human PRO307 prote

12	2044	96.8	383	22	Human Zsig13 varia
13	2037	95.4	383	22	Human membrane or
14	1873.5	88.7	375	21	Protein encoded by
15	1873.5	88.7	413	21	Protein encoded by
16	1868.5	88.5	382	21	Murine TANGO 186 p
17	1028	48.7	413	20	Human PRO1057 prot
18	1028	48.7	413	20	Human nn320_2 secr
19	1028	48.7	413	21	Human PRO1057 (UNQ
20	190.5	9.0	314	22	Bacillus lichenifo
21	185	8.8	316	13	Protease BLase. S
22	184	8.7	316	22	Bacillus lichenifo
23	172	8.1	222	22	Bacillus lichenifo
24	171	8.1	222	22	Bacillus lichenifo
25	171	8.1	222	22	Bacillus lichenifo
26	171	8.1	222	22	Bacillus lichenifo
27	171	8.1	222	22	Bacillus lichenifo
28	171	8.1	222	22	Bacillus lichenifo
29	169	8.0	222	22	Bacillus lichenifo
30	169	8.0	222	22	Bacillus lichenifo
31	169	8.0	222	22	Bacillus lichenifo
32	169	8.0	222	22	Bacillus lichenifo
33	168	8.0	222	22	Bacillus lichenifo
34	168	8.0	222	22	Bacillus lichenifo
35	168	8.0	222	22	Bacillus lichenifo
36	168	8.0	222	22	Bacillus lichenifo
37	167	7.9	222	22	Bacillus lichenifo
38	167	7.9	222	22	Bacillus lichenifo
39	167	7.9	222	22	Bacillus lichenifo
40	167	7.9	222	22	Bacillus lichenifo
41	167	7.9	222	22	Bacillus lichenifo
42	167	7.9	222	22	Bacillus lichenifo
43	167	7.9	222	22	Bacillus lichenifo
44	167	7.9	222	22	Bacillus lichenifo
45	167	7.9	222	22	Bacillus lichenifo

ALIGNMENTS

RESULT 1  
AAB48972  
ID AAB48972 standard; Protein; 392 AA.  
XX AAB48972;  
AC AAB48972;  
DT 27-MAR-2001 (first entry)  
XX  
DE Human Zsig13 variant #1, SEQ ID NO:2.  
XX

Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue;  
glutamy endopeptidase homologue; factor X homologue; trypsin homologue;  
trypsinogen homologue; mast cell protease homologue;  
collagenase homologue; protein degradation; food processing; brewing;  
alcohol production; laundry detergent component.  
Homo sapiens.  
US6153420-A.  
28-NOV-2000.  
04-MAY-1998; 98US-0072384.  
24-APR-1997; 97US-0044185.  
17-APR-1998; 98US-0062142.  
(ZYMO ) ZYMOGENETICS INC.  
Sheppard PO; - inventor  
WPI; 2001-060090/07.  
N-PSDB; AAC91782.

PT New isolated serine protease (designated Zsig13), useful in industrial  
 PT processes to degrade unwanted proteins or alter the characteristics of  
 PT protein-containing composition, as well as in industrial applications  
 PT (e.g. brewing)

PS Claim 1; Column 25-28; 26pp; English.

XX  
 CC The invention relates to human Zsig13 proteins (AAB48972-B48974), and  
 CC to DNA encoding them (AAC91782-C91784). The invention also relates to  
 CC expression vectors and host cells comprising a human Zsig13 DNA, and the  
 CC recombinant production of a human Zsig13 protein or its precursor.  
 CC Zsig13 is a serine protease, and has significant homology to Bacillus  
 CC licheniformis glutamyl endopeptidase, human clotting factor X, human  
 CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine  
 CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human  
 CC Zsig13 is located on chromosome 11q22.1. Zsig13 is useful in industrial  
 CC processes to degrade unwanted proteins or alter the characteristics of  
 CC protein-containing compositions. It may also be used in industrial  
 CC applications in which proteases are utilised, including food processing,  
 CC brewing and alcohol production, and as a component of a laundry  
 CC detergent. The present sequence represents a human Zsig13 variant.

XX Sequence 392 AA;

Query Match 99.6%; Score 2104; DB 22; Length 392;

Best Local Similarity 100.0%; Pred. No. 3.9e-147; Indels 0; Gaps 0;  
 Matches 392; Conservative 0; Mismatches 0;

QY 1 MAGIPGLLFLFLLCAGOVSPYSAPKPTWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60

DB 1 magipglilflilcavqgvspysapkwptwpayrlpvvlpqstlnlakpfgaeakle 60

QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSGDXGAXRDSGS 120

DB 61 vsscgpqchkgtpptykeakylysetlyangsrtextexvgvylsssgdgaaxrds 120

QY 121 SGKRRRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEXHVLTAACHIDG 180

DB 121 sgkrrrkqiygydsrfsifgkdfllnypfstsvklstgctglvaexhvltaachidg 180

QY 181 KTYVKGTKLRVGLPKFKFDGGRGANDSTAMPEQMKFQWIRKRVKTHVPKGWIKGNAND 240

DB 181 ktyvkgtklrvglpkfkfdgrgandstampeqmkfgwlrkvrkthvpkgwknand 240

QY 241 IGMDDYVALLLEKPKHKKRKMKGVSPPAKOLPGGRTHFSGYDNDRPNLVYRCDVKDE 300

DB 241 igmddyvallylekkphkrkkmkgvspakqlpggrthfsgydnrdpnlvyrfdv 300

QY 301 TYDLLYQCCDAQPGASGYGVYVVRMWRKQOQWERKTIIGIFSGHWDVMDNGSPQDFNVAVR 360

DB 301 tydlllyqccdaqpgasgygyvvrwmwrkqqkwerkiigifsgbhqwdvmdngspqdfnvavr 360

QY 361 ITPPKYAQICYWIKNYLDREGDTVFLPGSN 392

DB 361 itplkyaqicywiwnyldcregdtvflpgsn 392

RESULT 2

AAB48973

ID AAB48973 standard; Protein; 392 AA.

XX

AC AAB48973;

XX

DT 27-MAR-2001 (first entry)

XX Human Zsig13 variant #2, SEQ ID NO:15.

DE

KW Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue;  
 KW glutamyl endopeptidase homologue; factor X homologue; trypsin homologue;  
 KW trypsinogen homologue; mast cell protease homologue;  
 KW collagenase homologue; protein degradation; food processing; brewing;  
 KW alcohol production; laundry detergent component.

XX Homo sapiens.

XX US6153420-A.

XX

PD 28-NOV-2000.

XX

XX 04-MAY-1998; 98US-0072384.

XX

XX 24-APR-1997; 97US-0044185.

PR 17-APR-1998; 98US-0062142.

XX

PA (ZYMO ) ZYMOGENETICS INC.

XX

PI Sheppard PQ;

XX

DR WPI; 2001-060090/07.

DR

DR N-PSDB; AAC91783.

XX

PT New isolated serine protease (designated Zsig13), useful in industrial  
 PT processes to degrade unwanted proteins or alter the characteristics of  
 PT protein-containing composition, as well as in industrial applications  
 PT (e.g. brewing)

XX Claim 1; Column 35-38; 26pp; English.

XX The invention relates to human Zsig13 proteins (AAB48972-B48974), and  
 CC to DNA encoding them (AAC91782-C91784). The invention also relates to  
 CC expression vectors and host cells comprising a human Zsig13 DNA, and the  
 CC recombinant production of a human Zsig13 protein or its precursor.  
 CC Zsig13 is a serine protease, and has significant homology to Bacillus  
 CC licheniformis glutamyl endopeptidase, human clotting factor X, human  
 CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine  
 CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human  
 CC Zsig13 is located on chromosome 11q22.1. Zsig13 is useful in industrial  
 CC processes to degrade unwanted proteins or alter the characteristics of  
 CC protein-containing compositions. It may also be used in industrial  
 CC applications in which proteases are utilised, including food processing,  
 CC brewing and alcohol production, and as a component of a laundry  
 CC detergent. The present sequence represents a human Zsig13 variant.

XX Sequence 392 AA;

Query Match 98.7%; Score 2085; DB 22; Length 392;

Best Local Similarity 98.2%; Pred. No. 9.7e-146;

Matches 395; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGIPGLLFLFLLCAGOVSPYSAPKPTWPAYRLPVVLPSTLNLAKEPFGAEAKLE 60  
 DB 1 magipglilflilcavqgvspysapkwptwpayrlpvvlpqstlnlakpfgaeakle 60

QY 61 VSSCGPQCHKGTPPTYKEAKOYLSYETLYANGSRTEXOVGIYILSSGDXGAXRDSGS 120

DB 61 vsscgpqchkgtpptykeakylysetlyangsrtextexvgvylsssgdgaaxrds 120

QY 121 SGKRRRKQIYGYDSRFSIFGKDFLLNYPFSTSVKLTGCTGLVAEXHVLTAACHIDG 180

DB 121 sgkrrrkqiygydsrfsifgkdfllnypfstsvklstgctglvaexhvltaachidg 180

QY 181 KTYVKGTKLRVGLPKFKFDGGRGANDSTAMPEQMKFQWIRKRVKTHVPKGWIKGNAND 240

DB 181 ktyvkgtklrvglpkfkfdgrgandstampeqmkfgwlrkvrkthvpkgwknand 240

QY 241 IGMDDYVALLLEKPKHKKRKMKGVSPPAKOLPGGRTHFSGYDNDRPNLVYRCDVKDE 300

DB 241 igmddyvallylekkphkrkkmkgvspakqlpggrthfsgydnrdpnlvyrfdv 300

QY 301 TYDLLYQCCDAQPGASGYGVYVVRMWRKQOQWERKTIIGIFSGHWDVMDNGSPQDFNVAVR 360

DB 301 tydlllyqccdaqpgasgygyvvrwmwrkqqkwerkiigifsgbhqwdvmdngspqdfnvavr 360

QY 361 ITPPKYAQICYWIKNYLDREGDTVFLPGSN 392

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: December 2, 2001, 16:15:41 ; Search time 32.13 seconds  
(without alignments)  
447.327 Million cell updates/sec

Title: us-09-072-384-2

Perfect score: 2112

Sequence: 1 MAGIPGLFLFLFLCAVGG.....IKGNLYDCREGDTVFLPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	185	8.8	316	1	GSEP_BACLI
2	130.5	6.2	313	1	MPR_BACSU
3	122.5	5.8	490	1	FA10_RABIT
4	119.5	5.7	376	1	FA10_TROCA
5	118	5.6	266	1	EL1_BOVIN
6	118	5.6	269	1	EL2A_HUMAN
7	114	5.4	488	1	FA10_HUMAN
8	114	5.4	492	1	FA10_BOVIN
9	113	5.4	269	1	EL2_PIG
10	113	5.4	271	1	EL2_MOUSE
11	112.5	5.3	259	1	TRYP_STRGR
12	112	5.3	266	1	EL1_PIG
13	112	5.3	266	1	EL1_RAT
14	111.5	5.3	238	1	TRY5_AEDAE
15	111.5	5.3	686	1	MAS2_HUMAN
16	110	5.2	269	1	EL2B_HUMAN
17	109.5	5.2	875	1	NETR_HUMAN
18	109	5.2	253	1	CFAD_HUMAN
19	109	5.2	761	1	NETR_MOUSE
20	107	5.1	269	1	EL2_BOVIN
21	106	5.0	271	1	EL2_RAT
22	105	5.0	603	1	CFAL_MOUSE
23	104.5	4.9	268	1	TRYP_STRGA
24	104.5	4.9	273	1	YDGD_ECOLI
25	102	4.8	786	1	STUB_DROME
26	101.5	4.8	583	1	CFAL_HUMAN
27	99.5	4.7	281	1	TRY2_DROER
28	99.5	4.7	454	1	TMS3_HUMAN
29	99	4.7	274	1	TRY5_ANOGA
30	99	4.7	604	1	CFAL_RAT
31	97.5	4.6	259	1	DEF3_DERFA
32	96	4.5	407	1	FA7_BOVIN
33	95.5	4.5	261	1	DER3_DERPT

## RESULT 1

## GSEP\_BACLI

## ID

## AC

## P80057

## DT

## 01-MAY-1992

## Rel. 22, Created)

## DT

## 01-OCT-1993

## Rel. 27, Last sequence update)

## DT

## 20-AUG-2001

## Rel. 40, Last annotation update)

## DE

## GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC

## DE

## ENDOPEPTIDASE) (GSE).

## GN

## BLASE

## OS

## Bacillus licheniformis.

## OC

## Bacteria; Firmicutes; Bacillus/Clostridium group;

## OC

## Bacillus/Staphylococcus group; Bacillus.

## OX

## NCBI\_TaxID=1402;

## RN

## [1]

## RP

## SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

## RC

## STRAIN=ATCC 14580;

## RX

## MEDLINE=93054737; PubMed=1429718;

## RA

## Okamoto S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,

## RA

## Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;

## RT

## "Purification, characterization, cloning, and expression of a

## RT

## glutamic acid-specific protease from Bacillus licheniformis ATCC

## RT

## 14580.";

## RL

## J. Biol. Chem. 267:23782-23788(1992).

## RN

## [2]

## SEQUENCE OF 95-316.

## RP

## MEDLINE=92155199; PubMed=1346764;

## RX

## Svendsen I., Bredam K.;

## RA

## "Isolation and amino acid sequence of a glutamic acid specific

## RT

## endopeptidase from Bacillus licheniformis.";

## Eur. J. Biochem. 204:165-171(1992).

## CC

## FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE

## CC

## CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG

## CC

## PREFERENCE FOR GLU.

## CC

## CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-I-XAA, GLU-I-XAA.

## CC

## SUBCELLULAR LOCATION: SECRETED.

## CC

## SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE

## CC

## V8 FAMILY.

## CC

## This SWISS-PROT entry is copyright. It is produced through a collaboration

## CC

## between the Swiss Institute of Bioinformatics and the EMBL outstation -

## CC

## the European Bioinformatics Institute. There are no restrictions on its

## CC

## use by non-profit institutions as long as its content is in no way

## CC

## modified and this statement is not removed. Usage by and for commercial

## CC

## entities requires a license agreement (see http://www.isb-sib.ch/announce/

## CC

## or send an email to license@isb-sib.ch).

## CC

## EMBL; D10060; BAA00949.1; -

## DR

## PIR; S23078; S23078.

## DR

## PIR; A45134; A45134.

## DR

## MEROPS; S01.271; -

## DR

## InterPro; IPR00126; Ser\_proteas\_v8.

## DR

## InterPro; IPR001254; Trypsin.

## DR

## Pfam; PF00089; trypsin; 1.

## DR

## PRINTS; PR000839; V8PROTEASE.

## DR

## SMART; SM00020; TRYP\_SPC; 1.

Q9Y5Y6 homo sapien  
P56677 mus musculus  
P00741 bos taurus  
Q01513 podospora a  
P19540 canis famli  
O97370 euroglyphus  
Q99895 homo sapien  
P05049 drosophila  
P24664 saccharopol  
P52905 drosophila  
P08897 hypoderma l  
P21844 mus musculus

## ALIGNMENTS

RESULT 1  
GSEP\_BACLI  
ID GSEP\_BACLI STANDARD; PRT; 316 AA.  
AC P80057;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC

DE ENDOPEPTIDASE) (GSE).

GN BLASE.

OS Bacillus licheniformis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=1402;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=ATCC 14580;

RX MEDLINE=93054737; PubMed=1429718;

RA Okamoto S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,

RA Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;

RT "Purification, characterization, cloning, and expression of a

RT glutamic acid-specific protease from Bacillus licheniformis ATCC

RT 14580.";

RL J. Biol. Chem. 267:23782-23788(1992).

RN [2]

SEQUENCE OF 95-316.

RP MEDLINE=92155199; PubMed=1346764;

RX Svendsen I., Bredam K.;

RA "Isolation and amino acid sequence of a glutamic acid specific

RT endopeptidase from Bacillus licheniformis.";

RT Eur. J. Biochem. 204:165-171(1992).

CC FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE

CC CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG

CC PREFERENCE FOR GLU.

CC CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-I-XAA, GLU-I-XAA.

CC SUBCELLULAR LOCATION: SECRETED.

CC SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE

CC V8 FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

-----  
EMBL; D10060; BAA00949.1; -

DR PIR; S23078; S23078.

DR PIR; A45134; A45134.

DR MEROPS; S01.271; -

DR InterPro; IPR00126; Ser\_proteas\_v8.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR000839; V8PROTEASE.

DR SMART; SM00020; TRYP\_SPC; 1.

```
DR PROSITE; PS00672; V8_HIS; 1.  
DR PROSITE; PS00673; V8_SER; 1.  
KW Hydrolase; Serine protease; Signal.  
FT SIGNAL 1 ?  
FT PROPEP 2 94  
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.  
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 126 142  
FT DISULFID 275 279  
SQ SEQUENCE 316 AA; 96D7552CB7089B09 CRC64;  
  
Query Match 8.8%; Score 185; DB 1; Length 316;  
Best Local Similarity 23.2%; Pred. No. 6.6e-09;  
Matches 77; Conservative 46; Mismatches 139; Indels 70; Gaps 16;  
  
QY 63 SSCGPOCHKGTPL---PTYEAKQYLYETLYANGSTEXQVGIYILSSSGDGAXXRDSG 119  
DB 28 AQAAPSPH---TPVSSDPSYK-AETSVYDP-----NIKSDQYGLYSKAFGTG---KVNE 76  
QY 120 SSGKSRK-----RQIVGDSRFSIFGKDFLLNPFSTSVKLST---GCTGTILVAE 167  
DB 77 TKEAKRSKAPAKYISKVIGSDDRVTN---TTAYPRAIIVHSSIGSCCTGWMIGP 133  
QY 168 XHVLTAACHIDGKT-YVGTQKLRVGLPKPKFGDGRGANDSTSAMPEQMKFQWIRKVR 226  
DB 134 KTVATAGCHIVDTSSGSFAGTAVSPG-----RNGTFS-----YPIGSVKSTR 175  
QY 227 THVPKGIKNANDIGMDYDVALLEKPKHKKRPMKTCVSPPAKQLPGRHIFSGYDNR 286  
DB 176 YFIPSGWRSGNTN-----YDGAIEELSEPIGNTVGYFGSYTTSSLVGTGTVTISGYPGDK 230  
QY 287 PGNLVYRFCD-VKDEYDLYOQDAQGASGVGVVVRMVKRQOQKWKRII---GIF 340  
DB 231 TAGTQWHSFGPIAISEYKQLQYAM-DYGGQSGSPVEQSSRTNCSPGSLAVHTNGVY 289  
QY 341 SGHWDVMSGPDQNVAVRITPLKYAICVW 372  
DB 290 GG-----SSYNGRGTTRITREVFONLTNW 311  
  
RESULT 2  
MPR_BACSU STANDARD; PRT; 313 AA.  
AC P39790;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE EXTRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).  
GN MPR.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=GP241;  
RX MEDLINE=90130256; PubMed=2105291;  
RA Sloma A., Rudolph C.F., Rufo G.A. Jr., Sullivan B.J., Theriault K.A.,  
RA Ally D., Pero J.;  
RT "Gene encoding a novel extracellular metalloprotease in Bacillus  
RT subtilis.";  
RL J. Bacteriol. 172:1024-1029(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;  
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the  
RT Bacillus subtilis chromosome.";  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-68 FROM N.A.
```

```
RX MEDLINE=89108019; PubMed=3145906;  
RA Smith H., de Jong A., Bron S., Venema G.;  
RT "Characterization of signal-sequence-coding regions selected from the  
RT Bacillus subtilis chromosome.";  
RL Gene 70:351-361(1988).  
RN [4]  
RN SEQUENCE OF 1-10 FROM N.A.  
RP STRAIN=168;  
RX MEDLINE=96118702; PubMed=7496533;  
RA Saxild H.H., Jacobsen J.H., Nygaard P.;  
RT "Functional analysis of the Bacillus subtilis purt gene encoding  
RT formate-dependent glycinamide ribonucleotide transformylase.";  
RL Microbiology 141:2211-2218(1995).  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE  
CC V8 FAMILY.  
CC -1- CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2  
CC FAMILY OF SERINE PROTEASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
DR EMBL; L10505; AAA22604.1; -;  
DR EMBL; AB006424; BAA33121.1; -;  
DR EMBL; M22916; AAA22832.1; -;  
DR EMBL; Z99105; CAB12018.1; -;  
DR PIR; A35122; A35122.  
DR HSP; P00756; 1SGF.  
DR MEROPS; S01.272; -;  
DR Subtilist; BGI0690; mpr.  
DR InterPro; IPR000126; Ser_proteas_V8.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00839; V8PROTEASE.  
DR SMART; SM0020; Tryp_SPC; 1.  
DR PROSITE; PS00672; V8_HIS; 1.  
DR PROSITE; PS00673; V8_SER; 1.  
KW Hydrolase; Serine protease; Signal; Zymogen; Complete proteome.  
FT SIGNAL 1 34 POTENTIAL.  
FT PROPEP 35 93  
FT CHAIN 94 313 EXTRACELLULAR METALLOPROTEASE.  
FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 131 147 BY SIMILARITY.  
FT CONFLICT 61 68 QVSAPYEG -> PLESTAQA (IN REF. 3).  
SQ SEQUENCE 313 AA; 33842 MW; D41788ED652AE94 CRC64;  
  
Query Match 6.2%; Score 130.5; DB 1; Length 313;  
Best Local Similarity 22.6%; Pred. No. 0.00038;  
Matches 87; Conservative 38; Mismatches 139; Indels 121; Gaps 19;  
  
QY 28 WKTWPAYRLPVVLPOSTNLAKPDFCAEAKLEVS-----SSCGPOCHKGTPLTYEAK 82  
DB 7 FRQWFAYLIVLCALA---AAVSGVPAKAENPQTSVNTGKEA-DATKNQTSKADQ 61  
QY 83 QYLSYE-----TLYANGSRTEQVGIYILSSSGDGAXXRDSGSKSRKRQIYGYDS 135  
DB 62 VSAPYEGTGKTSKSLY--GGQTELEKNIQTLQPS-----SIIGTDE 100  
QY 136 RFSIFGKDFLLNYPFSTSVKLST-----GCTGTILVAEXHVLTAACHIH----- 178  
DB 101 RTRI---SSTSPFYRTVQLSLKYPNTSSTYCTGFLVNPNTVTVTAGHCVYSDHGWAS 157  
QY 179 -----DCKTVVKGQKLRVGLPKPKFGDGRGANDSTSAMPEQMKFQWIRKTRHP 230  
DB 158 TITAAPCRNGSSYPYGYI-----SGTMFYSVK-QWTESKDTNYD 195
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:14:45 ; Search time 48.03 Seconds  
(without alignments)  
621.704 Million cell updates/sec

Title: US-09-072-384-2

Perfect score: 2112

Sequence: 1 MAGTGPLLFLFLCAVQ.....IKGYLDCRGDTVFLPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68.\*

1: p1r1.\*

2: p1r2.\*

3: p1r3.\*

4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	185	8.8	316	2 A45134	endopeptidase (EC
2	130.5	6.2	313	2 A35122	metalloproteinase
3	125	5.9	482	1 EXRT	coagulation factor
4	118	5.6	269	2 B26823	pancreatic elastas
5	114	5.4	258	2 S70439	pancreatic elastasi
6	114	5.4	267	4 A56615	probable pancreati
7	114	5.4	488	1 EXHU	coagulation factor
8	114	5.4	492	1 EXBO	coagulation factor
9	113	5.4	269	2 A26823	pancreatic elastasi
10	113	5.4	271	2 A25528	pancreatic elastasi
11	112.5	5.3	259	1 TRSMG	trypsin (EC 3.4.21
12	112	5.3	266	1 ELRT1	pancreatic elastasi
13	112	5.3	266	1 ELPG	pancreatic elastasi
14	112	5.3	522	2 T29767	pancreatic elastasi
15	111.5	5.3	238	1 TRWV5V	hypothetical prote
16	111.5	5.3	686	1 A59271	trypsin-like prote
17	111	5.3	246	1 DBHU	Ra-reactive factor
18	110	5.2	269	2 C36823	complement factor
19	109	5.2	761	2 JC5759	pancreatic elastasi
20	107.5	5.1	405	2 T35117	brain-specific ser
21	106	5.0	236	2 A28566	probable secreted
22	106	5.0	271	1 ELRT2	T-cell suppressor
23	104.5	4.9	273	2 E85765	pancreatic elastasi
24	104.5	4.9	273	2 H64915	hypothetical prote
25	104.5	4.9	1582	2 T15308	hypothetical prote
26	102.5	4.9	2055	2 T31617	hypothetical prote
27	102	4.8	786	1 A47547	hypothetical prote
28	101.5	4.8	583	2 A29134	serine proteinase
29	101.5	4.8	747	2 I51579	complement factor

#### ALIGNMENTS

RESULT 1

A45134

endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis

C:Species: Bacillus licheniformis

C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999

C:Accession: A45134; S23078

R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S

J. Biol. Chem. 267, 23782-23788, 1992

A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s

A:Reference number: A45134; MUID:93054737

A:Accession: A45134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <KAK>

A:Cross-references: GB:D10060; MID:g216263; PIDN:BAA00949.1; PID:dl001415; PID:g21626

A:Experimental source: ATCC 14580

A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIP:118785)

R:Svensden, I.; Breddam, K.

Eur. J. Biochem. 204, 165-171, 1992

A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase.

A:Reference number: S23078; MUID:92155199

A:Accession: S23078

A:Status: preliminary

A:Molecule type: protein

A:Residues: 95-316 <SVE>

C:Keywords: hydrolase

Query Match 8.8%; Score 185; DB 2; Length 316;  
Best Local Similarity 23.2%; Pred. No. 2.9e-08;  
Matches 77; Conservative 46; Mismatches 139; Indels 70; Gaps 16;

QY 63 SCGQCQCHGTPL---PTYPEAKQYLSYETLYANGSRTEQVGIYLSSSGGGAXXRDSG 119

DB 28 AQAASPH--TPVSSDPSTYK-AETSVTDP-----NIKSDQGLYSKAFATGTG---KVNE 76

QY 120 SSGKSRK-----ROIYCYDSRFIFGKDFLLNYPFSTSVKLST---GCTGFLVAE 167

DB 77 TREKAEKSPAKAPYSIKSVIGSDDRTRVTN---TTAYPIRAIVHLSSIGSGCTGMNIGP 133

QY 168 XHVLTAACHIDGKT-YVKGTKLRVGLKPKFKDGGGRGANDSTAMPQMKFQWIRVR 226

DB 134 KTVATAGHCIIYDSSGSGFAGTATVSPG-----RNGTS-----YPGSVKSTR 175

QY 227 THVPKGTGNANDIGMDYDVALLELKKPKHKFKMKGIVSPPAKOLPGRIHFSGYDNR 286

DB 176 YFIPSGWRSGNTN-----YDYGAIELSEPIGNTVGVFGYSYTTSSLVGTTVTISGYPGDK 230

QY 287 FGNLYVRCD--VKDETDLLYQQDAQPGASGYGVYVPMKROQKWKERKII----GIF 340

DB 231 TAGTQWQHSPIAISETVKLOQYAM-DTYGGQSGSPVFQSSSRNTNCSPGCSLAVHTNGVY 289

masquerade precurs  
trypsin-related pr  
allergen Der f II  
trypsin-like prote  
trypsin-like prote  
coagulation factor  
probable secreted  
coagulation factor  
hypothetical prote  
adenylate cyclase  
trypsin-like prote  
probable polygalac  
coagulation factor  
probable trypsin V  
pancreatic elastasi  
serine proteinase  
chymase (EC 3.4.21

QY 341 SGHWDVNGSPQDFNVAVRITPLKYAICYW 372  
Db 290 GG-----SSYNGRTRITREVFDFNLTNW 311

RESULT 2  
A35122  
metalloprotease (EC 3.4.-.-) mpr precursor, extracellular - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 15-Oct-1999  
C:Accession: A35122; I40010; A69660  
R:Slona, A.; Rudolph, C.F.; Rufo Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; F  
J. Bacteriol. 172, 1024-1029, 1990  
A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.  
A:Reference number: A35122; MUID:90130256  
A:Accession: A35122  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-313 <SLO>  
A:Cross-references: GB:I10505; NID:gl43209; PIDN:AAA22604.1; PID:gl43210; GB:M29036  
R:Smith, R.; de Jong, A.; Bron, S.; Venema, G.  
Gene 70, 351-361, 1988  
A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s  
A:Reference number: I39994; MUID:89108019  
A:Accession: I40010  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-60, 65, 'L', 67, 'S', 69, 'AQA' <RES>  
A:Cross-references: GB:M22916; NID:gl43701; PIDN:AAA22832.1; PID:gl43702  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.; Bron, S.; Brouillet, S.; Bruchsi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maues  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudeg, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaite, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toononi, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69660  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-313 <KUN>  
A:Cross-references: GB:Z99105; GB:AL009126; NID:g2632457; PIDN:CAB12018.1; PID:el182176;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: mpr  
C:Keywords: hydrolase

Query Match 6.2%; Score 130.5; DB 2; Length 313;  
Best Local Similarity 22.6%; Pred. No. 0.0013;  
Matches 87; Conservative 38; Mismatches 139; Indels 121; Gaps 19;

QY 28 WKPTWPAAYRLPVVLPOSTNLNRPDCAEAKLEVS-----SSGQPOCHKGTPLPTTYEAK 82  
Db 7 FRKWFAYLTVLCALA-----AAVSGVPAKAAENPOTSVSNTGKEA-DRTKNOTSKADQ 61

QY 83 QYLSYE-----TLYANGSRTEQVGIYILSSGGDGAAXRDSGSSKRRKRIQVYDS 135  
Db 62 VSAPYEGTGTSKSLY--GGQTELEKNIQTLQPS-----SIIGTDE 100

QY 136 RFSIFGKDFLLNYPFSTVSKLST-----GCTGTLVAEXHVLTAACHIH 178  
Db 101 RTRI---SSTTSFPYRATVQLSIPKYPNTSSTYGTCTGLPNPNTVVTAGCHVCSODHGWS 157  
QY 179 -----DGKTYVKQKLRVGLFKPKFDGGGRANDSTSAMPEQMKFQWIRKRVHP 230

Db 158 TITAAPGRNGSSYPYGY-----SCTMEYSVK-GWTESKDNYD 195  
QY 231 KGWIKGN---ANDIGMDYDVALLELKKPKKMKIGVSPPAKQLPGGRITHFSGYDNDRP 287  
Db 196 YGAIKLNGSPONTVGW-YGYRTNSSSP-----VGLSSVTFGFCDKTFGTWSDTKP 247  
QY 288 GNLYRFCDVDETDYLLYOQDAQPGASGYGVYVVMKRRQOQKWKRIIGFSGHOWVD 347  
Db 248 IR-----SAETKYKLT-TTDTYGCQSGSPVY-----RNYSDTGQTATAIHT----- 287  
QY 348 MNGSPQDFNVAVRITPLKYAICYW 372  
Db 288 -NGG-SSYNLGTRVTDVFNNTQW 310

RESULT 3  
EXRT  
coagulation factor Xa (EC 3.4.21.6) precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Jan-1995 #sequence\_revision 07-Feb-1997 #text\_change 08-Dec-2000  
C:Accession: S49075; JC4670; PS0191; PS0190; I62745  
R:Stanton, C.; Ross, P.; Hutson, S.; Wallin, R.  
Thromb. Res. 80, 63-73, 1995  
A:Title: Evidence for competition between vitamin K-dependent clotting factors for in  
A:Reference number: A58498; MUID:96093366  
A:Accession: S49075  
A:Molecule type: mRNA  
A:Residues: 1-482 <STAI>  
A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
A:Note: Submitted to the EMBL Data Library, June 1994  
A:Note: neither the complete nucleic acid sequence nor the complete translation are s  
R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.  
Gene 169, 269-273, 1996  
A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.  
A:Reference number: JC4670; MUID:96194815  
A:Accession: JC4670  
A:Molecule type: mRNA  
A:Residues: 1-482 <STA2>  
A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
A:Experimental source: Cos-1 cell  
R:Enyoji, K.; Miyazaki, K.; Kato, H.  
J. Biochem. 109, 890-898, 1991  
A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat  
A:Reference number: PS0190; MUID:92041742  
A:Accession: PS0191  
A:Molecule type: protein  
A:Residues: 41-58, 'X', 60-65 <ENJ1>  
A:Accession: PS0190  
A:Molecule type: protein  
A:Residues: 183-186, 'X', 188-207 <ENJ2>  
R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.  
Eur. J. Haematol. 52, 162-168, 1994  
A:Title: Analysis of the partial nucleotide sequences and deduced primary structures  
A:Reference number: I46196; MUID:94222160  
A:Accession: I62745  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 295-383, 'G', 385-455 <MUR>  
A:Cross-references: GB:D21215; NID:g415309; PIDN:BAA04756.1; PID:g455396  
C:Function:  
A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the  
A:Pathway: blood coagulation  
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol  
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-40/Domain: propeptide #status predicted <PRO>  
F:45-84/Domain: Gla domain homology <GLA>  
F:41-179/Product: coagulation factor X light chain #status predicted <LCH>  
F:90-121/Domain: EGF homology <EG1>  
F:129-164/Domain: EGF homology <EG2>  
F:183-482/Product: coagulation factor X heavy chain #status predicted <HCH>  
F:183-231/Domain: activation peptide #status predicted <APT>